

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:24:09 ; Search time 54 Seconds  
(without alignments)  
2267.728 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 919  
Sequence: 1 MSKPYLFANRSEMPVALAAY.....TGVEEGNPILKHFITGTPF 919

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

A\_Geneseq\_101002:.\*  
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	919	100.0	919	21	AAV95987
2	788	85.7	889	21	AAV95988
3	11	1.2	576	21	AAV95821
4	11	1.2	578	21	AAV95820
5	9	1.0	262	21	AAV70579
6	9	1.0	851	14	AAAR41333
7	9	1.0	851	16	AAAR72077
8	9	1.0	851	17	AAW03166
9	9	1.0	851	21	AAAB12374
10	9	1.0	851	22	AAAB19963

11	9	1.0	851	23	AAE14649
12	9	1.0	851	23	AAE15173
13	8	0.9	114	22	ABG25555
14	8	0.9	331	23	ABBA7876
15	8	0.9	371	22	ABG25558
16	8	0.9	474	22	ABG17710
17	8	0.9	582	19	AAW56318
18	8	0.9	1373	22	ABB62368
19	7	0.8	29	14	AAAR4425
20	7	0.8	29	14	AAAR4426
21	7	0.8	29	14	AAAR4427
22	7	0.8	29	14	AAAR4428
23	7	0.8	34	16	AAAR4559
24	7	0.8	34	19	AAW51936
25	7	0.8	39	23	AAU88768
26	7	0.8	41	7	AAW60714
27	7	0.8	41	22	AAW60715
28	7	0.8	42	7	AAW60715
29	7	0.8	51	21	AAV64841
30	7	0.8	51	22	AAU22430
31	7	0.8	62	22	AAW87002
32	7	0.8	64	20	AAV11419
33	7	0.8	64	22	AAW77071
34	7	0.8	69	19	AAW75000
35	7	0.8	69	21	AAW75000
36	7	0.8	69	21	AAW75000
37	7	0.8	69	21	AAW75000
38	7	0.8	70	19	AAW74874
39	7	0.8	79	23	AAW75000
40	7	0.8	82	22	AAW75000
41	7	0.8	82	22	AAW75000
42	7	0.8	88	21	AAW75000
43	7	0.8	95	17	AAW9286
44	7	0.8	95	18	AAW44113
45	7	0.8	102	21	AAW54054
46	7	0.8	104	21	AAW57029
47	7	0.8	106	21	AAW37047
48	7	0.8	106	21	AAW37047
49	7	0.8	106	21	AAW37047
50	7	0.8	106	22	AAW33109
51	7	0.8	106	22	AAW38625
52	7	0.8	107	21	AAW37028
53	7	0.8	108	21	AAW37028
54	7	0.8	111	21	AAW40996
55	7	0.8	112	22	AAW65036
56	7	0.8	114	21	AAW41041
57	7	0.8	125	22	AAW01896
58	7	0.8	129	23	AAW01441
59	7	0.8	130	23	AAW47564
60	7	0.8	131	21	AAW00735
61	7	0.8	131	21	AAW00735
62	7	0.8	134	21	AAW40580
63	7	0.8	136	22	AAW051859
64	7	0.8	136	22	AAW03607
65	7	0.8	136	22	AAW06373
66	7	0.8	136	23	AAW33366
67	7	0.8	143	18	AAW18163
68	7	0.8	148	22	AAW22609
69	7	0.8	150	22	AAW87666
70	7	0.8	151	22	AAU17564
71	7	0.8	151	22	ABG21654
72	7	0.8	154	21	AAW41040
73	7	0.8	156	21	AAW37046
74	7	0.8	156	21	AAW47990
75	7	0.8	156	21	AAW47519
76	7	0.8	156	22	AAU41076
77	7	0.8	157	21	AAW16420
78	7	0.8	157	21	AAW41061
79	7	0.8	159	19	AAW63034
80	7	0.8	159	21	AAW14325
81	7	0.8	159	21	AAW91593
82	7	0.8	162	11	AAW07326
83	7	0.8	162	22	AAU47747
			162	22	AAW65590

Human STAT2 proteol  
Human Stat2 (sigma  
Novel human diagno  
Listeria monocytog  
Novel human diagno  
Novel human diagno  
Infectious lairngog  
Drosophila melanog  
Mutant alpha-amyli  
Mutant alpha-amyli  
Mutant alpha-amyli  
Protease Nexin-1 N  
Human protease nex  
Insulin/Insulin-11  
Sequence of the fu  
Gene #5 associated  
Sequence of the fu  
Human 5' EST relat  
Human cardiovascu  
Human Immune/Hema  
Human 5' EST secre  
Human colon cancer  
Human secreted pro  
Human secreted pro  
Human secreted pro  
Human secreted pro  
Human ORFX protein  
Human nervous syst  
E. coli growth and  
Zea mays protein f  
Human mtute Inter  
Human Interferon t  
Human pancreatic c  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Human digestive sy  
Human colorectal c  
Arabidopsis thalia  
Arabidopsis thalia  
Zea mays protein f  
Gene #5 associated  
Zea mays protein f  
Human polypeptide  
Human ORFX protein  
Listeria monocytog  
Human secreted pro  
Human secreted pro  
Human ORFX ORF344  
Propionibacterium  
Novel human diagno  
Human polypeptide  
Human ORF2339 prot  
Mycobacterium tube  
Novel human diagno  
Novel human diagno  
Novel signal trans  
Novel human diagno  
Zea mays protein f  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Propionibacterium  
Arabidopsis thalia  
Zea mays protein f  
Mycobacterium tube  
Mycobacterium tube  
Human secreted pro  
Fusion of alpha am  
Propionibacterium  
Propionibacterium

84	7	0.8	163	18	AAW41812	N. tabacum strain
85	7	0.8	169	11	AAW07250	Protein specific t
86	7	0.8	170	21	AAW44967	zea mays protein f
87	7	0.8	180	22	AAU77350	Propionibacterium
88	7	0.8	187	21	AAW56274	Human secreted pro
89	7	0.8	189	22	AAU40254	Propionibacterium
90	7	0.8	192	21	AAW40892	zea mays protein f
91	7	0.8	194	21	AAW55688	Arabidopsis thalia
92	7	0.8	195	4	AAW30076	Sequence of bovine
93	7	0.8	195	11	AAW04541	Bovine interferon-
94	7	0.8	195	19	AAW73227	Bovine interferon-
95	7	0.8	195	20	AAW83898	Bovine interferon-
96	7	0.8	197	21	AAW55687	Arabidopsis thalia
97	7	0.8	199	21	AAW55686	Arabidopsis thalia
98	7	0.8	213	19	AAW80649	S. pneumoniae tran
99	7	0.8	214	22	AAU17391	Novel signal trans
100	7	0.8	215	22	AAU17391	P. syringae pv. de
101	7	0.8	217	23	AAW12600	Human DIRP polype
102	7	0.8	224	22	AAW60158	Novel human diagno
103	7	0.8	227	22	AAW12473	Novel human secret
104	7	0.8	240	22	AAW62405	Human MBSP9 polype
105	7	0.8	255	23	AAW72381	Murine protein iso
106	7	0.8	261	10	AAW90438	Protein containing
107	7	0.8	261	21	AAW40566	Human ORFX ORF330
108	7	0.8	261	21	AAW94235	Claudin homologue
109	7	0.8	261	21	AAW94332	Human PRO1572 (UNQ
110	7	0.8	261	21	AAW70675	Human stomach prot
111	7	0.8	261	22	AAU29201	Human PRO polypept
112	7	0.8	261	22	AAW07051	Human gene 1 encod
113	7	0.8	261	22	AAW07073	Human gene 1 encod
114	7	0.8	261	22	AAW87584	Human PRO1572. Ho
115	7	0.8	261	22	AAW66181	Protein of the inv
116	7	0.8	261	23	AAW65078	Human albumin fusi
117	7	0.8	261	23	AAW65079	Human albumin fusi
118	7	0.8	261	23	AAW65079	Protein encoded by
119	7	0.8	261	23	AAW65079	Mouse Claudin-18 (
120	7	0.8	276	22	AAW05368	Lactococcus lactis
121	7	0.8	276	22	AAW05368	Human gene 1 encod
122	7	0.8	281	13	AAW28599	Human gene 1 encod
123	7	0.8	281	13	AAW28599	Human gene 1 encod
124	7	0.8	281	13	AAW28599	Human gene 1 encod
125	7	0.8	281	13	AAW28599	Human gene 1 encod
126	7	0.8	281	13	AAW28599	Human gene 1 encod
127	7	0.8	281	13	AAW28599	Human gene 1 encod
128	7	0.8	281	13	AAW28599	Human gene 1 encod
129	7	0.8	281	13	AAW28599	Human gene 1 encod
130	7	0.8	281	13	AAW28599	Human gene 1 encod
131	7	0.8	281	13	AAW28599	Human gene 1 encod
132	7	0.8	281	13	AAW28599	Human gene 1 encod
133	7	0.8	281	13	AAW28599	Human gene 1 encod
134	7	0.8	281	13	AAW28599	Human gene 1 encod
135	7	0.8	281	13	AAW28599	Human gene 1 encod
136	7	0.8	281	13	AAW28599	Human gene 1 encod
137	7	0.8	281	13	AAW28599	Human gene 1 encod
138	7	0.8	281	13	AAW28599	Human gene 1 encod
139	7	0.8	281	13	AAW28599	Human gene 1 encod
140	7	0.8	281	13	AAW28599	Human gene 1 encod
141	7	0.8	281	13	AAW28599	Human gene 1 encod
142	7	0.8	281	13	AAW28599	Human gene 1 encod
143	7	0.8	281	13	AAW28599	Human gene 1 encod
144	7	0.8	281	13	AAW28599	Human gene 1 encod
145	7	0.8	281	13	AAW28599	Human gene 1 encod
146	7	0.8	281	13	AAW28599	Human gene 1 encod
147	7	0.8	281	13	AAW28599	Human gene 1 encod
148	7	0.8	281	13	AAW28599	Human gene 1 encod
149	7	0.8	281	13	AAW28599	Human gene 1 encod
150	7	0.8	281	13	AAW28599	Human gene 1 encod
151	7	0.8	281	13	AAW28599	Human gene 1 encod
152	7	0.8	281	13	AAW28599	Human gene 1 encod
153	7	0.8	281	13	AAW28599	Human gene 1 encod
154	7	0.8	281	13	AAW28599	Human gene 1 encod
155	7	0.8	281	13	AAW28599	Human gene 1 encod
156	7	0.8	281	13	AAW28599	Human gene 1 encod

230	7	0.8	398	23	ABR09439	Serpin domain prot
231	7	0.8	400	22	ABG02720	Novel human diapo
232	7	0.8	408	22	AA330705	A Bacillus pectate
233	7	0.8	412	20	AA355332	Chlamydia pneumoni
234	7	0.8	419	22	AA64997	Human secreted pro
235	7	0.8	420	23	ABR76313	Human protein phos
236	7	0.8	433	22	ABR67364	Drosophila melanog
237	7	0.8	433	23	ABR61037	Lactobacillus rham
238	7	0.8	434	17	AA966419	Peptide fragment o
239	7	0.8	434	21	AA12820	Rat N-type calcium
240	7	0.8	441	21	AA36179	Arabidopsis thalia
241	7	0.8	443	21	AA91592	Human secreted pro
242	7	0.8	443	22	ABG26155	Novel human diapo
243	7	0.8	443	23	ABR93726	Herbically activ
244	7	0.8	449	21	AA24869	Plant SDF encoded
245	7	0.8	458	23	AAU81986	Human secreted pro
246	7	0.8	468	22	ABG00728	Novel human diapo
247	7	0.8	469	22	AA40870	Human polypeptide
248	7	0.8	474	22	AA689794	C glutamicum prote
249	7	0.8	480	22	AAU33686	Pseudomonas aerugi
250	7	0.8	481	22	ABR70673	Drosophila melanog
251	7	0.8	488	19	AA49025	Kluyveromyces lact
252	7	0.8	491	22	AAU87389	Novel central nerv
253	7	0.8	491	22	AAU17172	Novel signal trans
254	7	0.8	491	22	AAU07844	Novel human serine
255	7	0.8	494	22	AA692620	C glutamicum prote
256	7	0.8	494	22	AA878986	C. glutamicum SPT
257	7	0.8	497	8	AA70572	Glucanase gene
258	7	0.8	511	8	AA70339	Alpha-amylose gene
259	7	0.8	511	19	AA39741	B. licheniformis a
260	7	0.8	511	20	AAU1582	An alpha-amylose p
261	7	0.8	511	20	AAU01583	An alpha-amylose p
262	7	0.8	512	8	AA70753	phea atrof amy oper
263	7	0.8	512	18	AAU10325	B. licheniformis a
264	7	0.8	512	20	AA722181	Wild type alpha-am
265	7	0.8	512	20	AAU00769	B. licheniformis a
266	7	0.8	512	20	AAW73509	Alpha-amylose prot
267	7	0.8	512	22	AA65875	B. licheniformis a
268	7	0.8	512	22	AA65877	B. licheniformis a
269	7	0.8	512	22	AAU07097	Bacillus lichenifo
270	7	0.8	512	22	AAU07098	Bacillus lichenifo
271	7	0.8	512	22	AAU07099	Bacillus lichenifo
272	7	0.8	512	22	AAU07100	Bacillus lichenifo
273	7	0.8	512	22	AAU07549	Bacillus lichenifo
274	7	0.8	512	22	AAU07550	Bacillus lichenifo
275	7	0.8	512	22	AAU07551	Bacillus lichenifo
276	7	0.8	512	22	AAU07552	Bacillus lichenifo
277	7	0.8	512	22	AAU07553	Bacillus lichenifo
278	7	0.8	512	22	AAU07554	Bacillus lichenifo
279	7	0.8	512	22	AAU07555	Bacillus lichenifo
280	7	0.8	512	22	AAU07556	Bacillus lichenifo
281	7	0.8	513	22	AAU31141	Novel human secret
282	7	0.8	516	20	AAW93433	A. thaliana ELV pr
283	7	0.8	517	18	AA34204	Streptomyces efllu
284	7	0.8	517	19	AAW58605	Streptomyces roseo
285	7	0.8	520	11	AAK04571	ORE3 product from
286	7	0.8	524	22	AA662989	Drosophila melanog
287	7	0.8	524	22	ABR62990	Drosophila melanog
288	7	0.8	534	20	AA729133	Amino acid sequenc
289	7	0.8	548	22	ABR62279	Drosophila melanog
290	7	0.8	551	21	AA16422	Pinus radiata 4CL
291	7	0.8	560	18	AAW23602	Alpha-amylose-cell
292	7	0.8	560	18	AAW22522	Alpha-amylose-cell
293	7	0.8	560	22	AA888599	Human hydrophobic
294	7	0.8	567	22	ABR68354	Human protein melano
295	7	0.8	567	22	AA894236	Human protein sequ
296	7	0.8	574	19	AAW57434	Teramyl-1-linker-CB
297	7	0.8	593	23	AA330097	Streptococcus poly
298	7	0.8	597	22	ABG18227	Novel human diapo
299	7	0.8	599	22	AA691428	C glutamicum prote
300	7	0.8	599	22	AA876686	Corynebacterium gl
301	7	0.8	624	17	AA877674	Glucanase from
302	7	0.8	630	18	AAW23603	Alpha-amylose-Cena
303	7	0.8	630	18	AAW22523	Alpha-amylose-Cena
304	7	0.8	632	21	AA801385	Neuron-associated
305	7	0.8	632	21	AA66689	Membrane-bound pro
306	7	0.8	632	22	AAU29106	Human PPO polypept
307	7	0.8	632	22	AAW39084	Human polypeptide
308	7	0.8	632	22	AA887545	Human PRO1136. HO
309	7	0.8	632	22	AA865212	Human PRO1136 (UNQ
310	7	0.8	640	22	AA678650	Plasmodium regula
311	7	0.8	640	22	ABG03838	Novel human diapo
312	7	0.8	653	22	ABR57870	Drosophila melanog
313	7	0.8	655	21	AA825102	Pinus radiata cell
314	7	0.8	658	21	AA833793	S. fradiae tylosin
315	7	0.8	650	23	AAE20142	Human protein cont
316	7	0.8	694	22	ABR70601	Drosophila melanog
317	7	0.8	696	22	AAU17394	Novel signal trans
318	7	0.8	704	20	AA531754	Candida tropicalis
319	7	0.8	706	22	ABR65428	Drosophila melanog
320	7	0.8	709	21	AA914466	Human secreted pro
321	7	0.8	719	22	AAW52608	Aquifex pyrophilus
322	7	0.8	723	22	ABG29456	Novel human diapo
323	7	0.8	749	23	AAW49640	Human tumour-assoc
324	7	0.8	750	18	AAW23600	Alpha-amylose-cell
325	7	0.8	750	18	AAW22520	Alpha-amylose-cell
326	7	0.8	782	19	AAW57436	Teramyl-1-linker-CB
327	7	0.8	786	15	AAW47066	Sequence of Crypto
328	7	0.8	802	22	ABR71892	Drosophila melanog
329	7	0.8	815	23	ABR1806	Herbically activ
330	7	0.8	819	23	AA69420	Lung small cell ca
331	7	0.8	823	22	ABG09169	Novel human diapo
332	7	0.8	823	22	ABG26009	Novel human diapo
333	7	0.8	834	22	AA830701	A Bacillus pectate
334	7	0.8	836	23	AAW49641	Human tumour-assoc
335	7	0.8	857	21	AA869355	HIV-1 non-subtype
336	7	0.8	984	21	AA825544	Eucalyptus grandis
337	7	0.8	999	21	AA653685	Arabidopsis thalia
338	7	0.8	1008	21	AA653684	Arabidopsis thalia
339	7	0.8	1014	21	AA825488	Eucalyptus grandis
340	7	0.8	1028	22	ABR62462	Drosophila melanog
341	7	0.8	1031	22	ABR65127	Drosophila melanog
342	7	0.8	1049	19	AAW73062	Rat GRP2. Ratius
343	7	0.8	1065	21	AA653683	Arabidopsis thalia
344	7	0.8	1068	14	AA833341	p110. AAR43341;
345	7	0.8	1068	14	AA833342	Human p110. Homo
346	7	0.8	1068	15	AA846294	Ptdins 3-kinase 11
347	7	0.8	1068	23	AAU09687	Human p110alpha is
348	7	0.8	1090	22	AA895857	Human protein sequ
349	7	0.8	1097	22	ABR11833	Human Cl channel n
350	7	0.8	1132	22	ABR65468	Drosophila melanog
351	7	0.8	1128	17	AAW77673	S-layer protein en
352	7	0.8	1228	18	AAW22862	Bacillus stearothe
353	7	0.8	1228	20	AAW93252	B. stearothermophi
354	7	0.8	1228	21	AA810625	B. stearothermophi
355	7	0.8	1229	22	ABG19475	Novel human diapo
356	7	0.8	1268	23	AAU10650	Chicken Nr-CAM pro
357	7	0.8	1326	22	ABR61156	Drosophila melanog
358	7	0.8	1335	22	ABR71593	Drosophila melanog
359	7	0.8	1394	22	ABG22611	Novel human diapo
360	7	0.8	1416	22	ABR65499	Drosophila melanog
361	7	0.8	1434	21	AA852028	P. horikoshii PHBN0
362	7	0.8	1454	21	AA851657	P. horikoshii PHB
363	7	0.8	1643	22	ABR63371	Drosophila melanog
364	7	0.8	2153	22	AAU33195	Novel human secret
365	7	0.8	2969	22	AA690600	C glutamicum prote
366	6	0.7	8	21	AA691117	Subsequence which
367	6	0.7	9	15	AA838144	Hepatitis B virus-
368	6	0.7	9	15	AA859185	Hepatitis B virus-
369	6	0.7	9	18	AAW39554	HBV pol peptide (p
370	6	0.7	9	19	AAW78884	Hepatitis B virus
371	6	0.7	9	19	AAW54507	Synthetic polypept
372	6	0.7	9	20	AA4545712	Immunogenic peptid
373	6	0.7	9	20	AA446774	Immunogenic peptid
374	6	0.7	9	21	AA833696	MHC class I assoc
375	6	0.7	9	21	AA806912	Claudin cell adhes

376	6	0.7	9	22	AA93792	Hepatitis B virus
377	6	0.7	9	22	AA95942	MHC class-I associ
378	6	0.7	9	22	AA97545	CAR sequence, a ca
379	6	0.7	10	15	AA73863	Antigen fragment 1
380	6	0.7	10	21	AA848614	CaMK I Ser/Thr kin
381	6	0.7	10	22	AAU06386	Human prostate ant
382	6	0.7	10	22	AAU06420	Human leukocyte an
383	6	0.7	10	22	AAU06421	Human leukocyte an
384	6	0.7	10	22	AAU06421	Human leukocyte an
385	6	0.7	10	22	AAU06421	Human complementar
386	6	0.7	10	22	AAU06421	Human complementar
387	6	0.7	11	21	AA79791	HIV-2 detection pho
388	6	0.7	11	21	AA79791	HIV-2 detection pho
389	6	0.7	11	21	AA79791	HIV-2 detection pho
390	6	0.7	11	21	AA79791	HIV-2 detection pho
391	6	0.7	11	21	AA79791	HIV-2 detection pho
392	6	0.7	11	21	AA79791	HIV-2 detection pho
393	6	0.7	11	21	AA79791	HIV-2 detection pho
394	6	0.7	11	21	AA79791	HIV-2 detection pho
395	6	0.7	11	21	AA79791	HIV-2 detection pho
396	6	0.7	11	21	AA79791	HIV-2 detection pho
397	6	0.7	11	21	AA79791	HIV-2 detection pho
398	6	0.7	11	21	AA79791	HIV-2 detection pho
399	6	0.7	11	21	AA79791	HIV-2 detection pho
400	6	0.7	11	21	AA79791	HIV-2 detection pho
401	6	0.7	11	21	AA79791	HIV-2 detection pho
402	6	0.7	11	21	AA79791	HIV-2 detection pho
403	6	0.7	11	21	AA79791	HIV-2 detection pho
404	6	0.7	11	21	AA79791	HIV-2 detection pho
405	6	0.7	11	21	AA79791	HIV-2 detection pho
406	6	0.7	11	21	AA79791	HIV-2 detection pho
407	6	0.7	11	21	AA79791	HIV-2 detection pho
408	6	0.7	11	21	AA79791	HIV-2 detection pho
409	6	0.7	11	21	AA79791	HIV-2 detection pho
410	6	0.7	11	21	AA79791	HIV-2 detection pho
411	6	0.7	11	21	AA79791	HIV-2 detection pho
412	6	0.7	11	21	AA79791	HIV-2 detection pho
413	6	0.7	11	21	AA79791	HIV-2 detection pho
414	6	0.7	11	21	AA79791	HIV-2 detection pho
415	6	0.7	11	21	AA79791	HIV-2 detection pho
416	6	0.7	11	21	AA79791	HIV-2 detection pho
417	6	0.7	11	21	AA79791	HIV-2 detection pho
418	6	0.7	11	21	AA79791	HIV-2 detection pho
419	6	0.7	11	21	AA79791	HIV-2 detection pho
420	6	0.7	11	21	AA79791	HIV-2 detection pho
421	6	0.7	11	21	AA79791	HIV-2 detection pho
422	6	0.7	11	21	AA79791	HIV-2 detection pho
423	6	0.7	11	21	AA79791	HIV-2 detection pho
424	6	0.7	11	21	AA79791	HIV-2 detection pho
425	6	0.7	11	21	AA79791	HIV-2 detection pho
426	6	0.7	11	21	AA79791	HIV-2 detection pho
427	6	0.7	11	21	AA79791	HIV-2 detection pho
428	6	0.7	11	21	AA79791	HIV-2 detection pho
429	6	0.7	11	21	AA79791	HIV-2 detection pho
430	6	0.7	11	21	AA79791	HIV-2 detection pho
431	6	0.7	11	21	AA79791	HIV-2 detection pho
432	6	0.7	11	21	AA79791	HIV-2 detection pho
433	6	0.7	11	21	AA79791	HIV-2 detection pho
434	6	0.7	11	21	AA79791	HIV-2 detection pho
435	6	0.7	11	21	AA79791	HIV-2 detection pho
436	6	0.7	11	21	AA79791	HIV-2 detection pho
437	6	0.7	11	21	AA79791	HIV-2 detection pho
438	6	0.7	11	21	AA79791	HIV-2 detection pho
439	6	0.7	11	21	AA79791	HIV-2 detection pho
440	6	0.7	11	21	AA79791	HIV-2 detection pho
441	6	0.7	11	21	AA79791	HIV-2 detection pho
442	6	0.7	11	21	AA79791	HIV-2 detection pho
443	6	0.7	11	21	AA79791	HIV-2 detection pho
444	6	0.7	11	21	AA79791	HIV-2 detection pho
445	6	0.7	11	21	AA79791	HIV-2 detection pho
446	6	0.7	11	21	AA79791	HIV-2 detection pho
447	6	0.7	11	21	AA79791	HIV-2 detection pho
448	6	0.7	11	21	AA79791	HIV-2 detection pho



522	6	0.7	39	23	AAU88751	Insulin/insulin-11	595	6	0.7	50	22	AAU33786	Peptide #7823 enco
523	6	0.7	39	23	AAU88761	Insulin/insulin-11	596	6	0.7	50	23	ABG43459	Human peptide enco
524	6	0.7	39	23	AAU88765	Insulin/insulin-11	597	6	0.7	50	23	AAU90802	Insulin/insulin-11
525	6	0.7	39	23	AAU88771	Insulin/insulin-11	598	6	0.7	52	21	AAU09489	Arabidopsis thalia
526	6	0.7	39	23	AAU88773	Insulin/insulin-11	599	6	0.7	52	21	AAU02982	Human secreted pro
527	6	0.7	39	23	AAU90800	Insulin/insulin-11	600	6	0.7	52	22	AAU51595	Propionibacterium
528	6	0.7	40	19	AAU78737	Leader peptide seq	601	6	0.7	53	21	AAU50046	Arabidopsis thalia
529	6	0.7	40	19	AAU78741	Leader peptide seq	602	6	0.7	54	21	AAU50045	Arabidopsis thalia
530	6	0.7	40	20	AAU72333	Xylanase II secret	603	6	0.7	54	21	AAU53245	Arabidopsis thalia
531	6	0.7	40	20	AAU12038	Human 5' EST secret	604	6	0.7	54	22	AAU39729	Propionibacterium
532	6	0.7	40	21	AAU58409	Lung cancer associ	605	6	0.7	54	22	AAU42051	Propionibacterium
533	6	0.7	40	21	AAU79833	HIV infection dete	606	6	0.7	54	22	AAU63984	Propionibacterium
534	6	0.7	40	22	ABU36101	Peptide #3607 enco	607	6	0.7	54	22	AAU66104	Propionibacterium
535	6	0.7	40	22	ABU21491	Protein #3490 enco	608	6	0.7	54	23	ABU51854	Human colon specif
536	6	0.7	40	22	AAU56882	Human brain expres	609	6	0.7	54	23	ABU50501	Human ORFX protein
537	6	0.7	40	22	AAU69272	Human bone marrow	610	6	0.7	55	22	AAU43640	Propionibacterium
538	6	0.7	40	22	AAU17102	Peptide #3536 enco	611	6	0.7	55	22	AAU47749	Propionibacterium
539	6	0.7	40	22	AAU29592	Peptide #3629 enco	612	6	0.7	55	22	AAU49783	Propionibacterium
540	6	0.7	40	22	AAU74059	Human colon cancer	613	6	0.7	55	22	AAU22774	Human prostate can
541	6	0.7	40	22	AAU74403	Vector pXBGI-TV en	614	6	0.7	55	22	AAU94755	Human reproductive
542	6	0.7	40	23	ABG38885	Human peptide enco	615	6	0.7	56	22	AAU04653	Haem protein relat
543	6	0.7	40	23	AAU88752	Insulin/insulin-11	616	6	0.7	57	22	AAU61850	Propionibacterium
544	6	0.7	41	16	AAU80538	Moloney murine leu	617	6	0.7	57	22	AAU63889	Propionibacterium
545	6	0.7	41	16	AAU76597	MOuV gag matrix p	618	6	0.7	57	22	AAU92691	Human digestive sy
546	6	0.7	41	17	AAU88178	Leader sequence fo	619	6	0.7	57	22	AAU23824	Human EST encoded
547	6	0.7	41	19	AAU78727	Leader peptide seq	620	6	0.7	58	22	AAU59603	Propionibacterium
548	6	0.7	41	19	AAU78735	Leader peptide seq	621	6	0.7	58	22	AAU59909	Propionibacterium
549	6	0.7	41	19	AAU78738	Leader peptide seq	622	6	0.7	58	22	AAU85462	Human immune/haema
550	6	0.7	41	19	AAU78739	Leader peptide seq	623	6	0.7	59	19	AAU76664	Elapidae modified
551	6	0.7	41	21	AAU56118	Human secreted pro	624	6	0.7	59	19	AAU76666	Elapidae modified
552	6	0.7	41	21	AAU27179	PIV3 L polymerase	625	6	0.7	59	21	AAU55754	Arabidopsis thalia
553	6	0.7	41	21	AAU27197	PIV3 L polymerase	626	6	0.7	59	21	AAU59560	Arabidopsis thalia
554	6	0.7	42	19	AAU21570	Cone snail alpha-c	627	6	0.7	59	22	AAU61825	Propionibacterium
555	6	0.7	42	19	AAU78719	Leader peptide seq	628	6	0.7	59	23	ABU33467	Human ORF2440 prot
556	6	0.7	42	19	AAU78722	Leader peptide seq	629	6	0.7	60	21	AAU15112	Arabidopsis thalia
557	6	0.7	42	21	AAU79774	HIV detection rela	630	6	0.7	60	22	AAU39247	Propionibacterium
558	6	0.7	42	21	AAU79840	HIV infection dete	631	6	0.7	60	22	AAU64705	Propionibacterium
559	6	0.7	42	22	ABU02387	Novel human diagno	632	6	0.7	61	21	AAU07969	Arabidopsis thalia
560	6	0.7	43	18	AAU19135	LA19 leader peptid	633	6	0.7	61	21	AAU12116	Arabidopsis thalia
561	6	0.7	43	19	AAU78726	Leader peptide seq	634	6	0.7	61	21	AAU00892	Human secreted pro
562	6	0.7	43	19	AAU78717	Leader peptide seq	635	6	0.7	61	22	ABU59394	Drosophila melanog
563	6	0.7	43	19	AAU78718	Leader peptide seq	636	6	0.7	61	22	AAU88413	Human immune/haema
564	6	0.7	43	19	AAU69152	Yeast expression s	637	6	0.7	62	22	ABU27472	Streptococcus poly
565	6	0.7	43	19	AAU69151	Yeast expression s	638	6	0.7	62	22	ABU43391	Peptide #10897 enc
566	6	0.7	43	20	AAU13987	P60 repeat sequenc	639	6	0.7	62	22	ABU16601	Human nervous syst
567	6	0.7	43	23	AAU90779	Insulin/insulin-11	640	6	0.7	62	22	ABU17995	Human nervous syst
568	6	0.7	44	22	AAU89297	Human immune/haema	641	6	0.7	62	22	ABU26365	Protein #8364 enco
569	6	0.7	44	22	AAU10101	Human polypeptide	642	6	0.7	62	22	AAU64315	Human brain expres
570	6	0.7	46	21	AAU79813	HIV infection dete	643	6	0.7	62	22	AAU77143	Human bone marrow
571	6	0.7	46	21	AAU79815	HIV infection dete	644	6	0.7	62	22	AAU83881	Human immune/haema
572	6	0.7	46	22	ABU31712	Peptide #4363 enco	645	6	0.7	62	22	AAU21077	Peptide #7511 enco
573	6	0.7	46	22	ABU36932	Peptide #4438 enco	646	6	0.7	62	22	AAU37272	Peptide #11309 enc
574	6	0.7	46	22	ABU40181	Peptide #7687 enco	647	6	0.7	62	23	ABU46154	Human peptide enco
575	6	0.7	46	22	ABU24627	Protein #6626 enco	648	6	0.7	62	23	ABU03006	Human ORFX protein
576	6	0.7	46	22	AAU60949	Human brain expres	649	6	0.7	63	22	AAU39445	Propionibacterium
577	6	0.7	46	22	AAU73642	Human bone marrow	650	6	0.7	63	22	AAU34679	Propionibacterium
578	6	0.7	46	22	AAU20047	Peptide #6481 enco	651	6	0.7	63	22	AAU51588	Propionibacterium
579	6	0.7	46	22	AAU30414	Peptide #4451 enco	652	6	0.7	63	23	ABU32712	Human cadherin-11k
580	6	0.7	46	22	AAU33833	Peptide #7870 enco	653	6	0.7	63	23	ABU11177	Human ORFX protein
581	6	0.7	46	23	ABU39710	Human peptide enco	654	6	0.7	64	21	AAU65012	Human 5' EST relat
582	6	0.7	46	23	ABU43514	Human peptide enco	655	6	0.7	64	22	AAU44953	Propionibacterium
583	6	0.7	47	22	AAU86352	Human immune/haema	656	6	0.7	64	22	AAU92782	Human digestive sy
584	6	0.7	47	22	AAU87232	Human immune/haema	657	6	0.7	64	22	AAU64709	Propionibacterium
585	6	0.7	48	20	AAU73915	Human prostate tum	658	6	0.7	65	22	AAU09187	Human polypeptide
586	6	0.7	48	22	AAU59509	Transactivation do	659	6	0.7	65	23	ABU04894	Human ORFX protein
587	6	0.7	49	16	AAU77796	Transactivation do	660	6	0.7	66	21	AAU54237	Human pancreatic c
588	6	0.7	49	16	AAU77797	Transactivation do	661	6	0.7	66	22	AAU43022	Propionibacterium
589	6	0.7	49	21	AAU34649	Arabidopsis thalia	662	6	0.7	66	22	AAU45044	Propionibacterium
590	6	0.7	50	22	AAU55386	Propionibacterium	663	6	0.7	66	22	AAU62876	Propionibacterium
591	6	0.7	50	22	ABU40144	Peptide #7650 enco	664	6	0.7	66	22	AAU31701	Novel human secret
592	6	0.7	50	22	AAU60907	Human brain expres	665	6	0.7	66	22	AAU07819	Human polypeptide
593	6	0.7	50	22	AAU73591	Human bone marrow	666	6	0.7	66	23	ABU00389	Human ORFX protein
594	6	0.7	50	22	AAU13825	Human polypeptide	667	6	0.7	66	23	ABU05956	Human ORFX protein

658	6	0.7	67	18	AAW27986	Staphylococcus aur
659	6	0.7	67	22	ABG23990	Novel human diagno
670	6	0.7	67	22	ABG29430	Novel human diagno
671	6	0.7	67	22	AAU20806	Human novel foetal
672	6	0.7	68	12	AAU20806	Kistlin. Syntheti
673	6	0.7	68	13	AAU20806	PIA from Agkistrod
674	6	0.7	68	15	AAU25191	Disintegrin peptid
675	6	0.7	68	21	AAU53943	M. genitalium ycfB
676	6	0.7	68	22	AAU55657	Human testicular a
677	6	0.7	68	22	AAU55657	Human reproductive
678	6	0.7	68	22	AAU55657	Human immune/haema
679	6	0.7	68	23	ABP38811	Staphylococcus epi
680	6	0.7	68	23	ABP38811	Human ORFX protein
681	6	0.7	69	21	AAU58275	Arabidopsis thalia
682	6	0.7	69	22	ABP06283	Drosophila melanog
683	6	0.7	69	22	ABP06283	Protonibacterium
684	6	0.7	69	22	AAU45102	Novel human diagno
685	6	0.7	69	22	ABG29142	Human peptide #941
686	6	0.7	69	22	ABG28290	Peptide #6265 enco
687	6	0.7	69	22	ABG28290	Protein #929 enco
688	6	0.7	69	22	ABG28290	Human brain expres
689	6	0.7	69	22	ABG28290	Human brain expres
690	6	0.7	69	22	ABG28290	Human bone marrow
691	6	0.7	69	22	AAU54247	Peptide #947 enco
692	6	0.7	69	22	AAU54247	Peptide #967 enco
693	6	0.7	69	22	AAU54247	Peptide #6269 enco
694	6	0.7	69	22	AAU54247	Peptide #921 enco
695	6	0.7	69	22	AAU54247	Human peptide enco
696	6	0.7	69	22	AAU54247	Human ORFX protein
697	6	0.7	69	22	AAU54247	Drosophila melanog
698	6	0.7	69	22	AAU54247	Protonibacterium
699	6	0.7	69	22	AAU54247	Human ORFX protein
700	6	0.7	69	22	AAU54247	Human bone marrow
701	6	0.7	69	22	AAU54247	Peptide #947 enco
702	6	0.7	69	22	AAU54247	Peptide #967 enco
703	6	0.7	69	22	AAU54247	Peptide #6269 enco
704	6	0.7	69	22	AAU54247	Peptide #921 enco
705	6	0.7	69	22	AAU54247	Human peptide enco
706	6	0.7	69	22	AAU54247	Human ORFX protein
707	6	0.7	69	22	AAU54247	Drosophila melanog
708	6	0.7	69	22	AAU54247	Protonibacterium
709	6	0.7	69	22	AAU54247	Human ORFX protein
710	6	0.7	69	22	AAU54247	Human bone marrow
711	6	0.7	69	22	AAU54247	Peptide #947 enco
712	6	0.7	69	22	AAU54247	Peptide #967 enco
713	6	0.7	69	22	AAU54247	Peptide #6269 enco
714	6	0.7	69	22	AAU54247	Peptide #921 enco
715	6	0.7	69	22	AAU54247	Human peptide enco
716	6	0.7	69	22	AAU54247	Human ORFX protein
717	6	0.7	69	22	AAU54247	Drosophila melanog
718	6	0.7	69	22	AAU54247	Protonibacterium
719	6	0.7	69	22	AAU54247	Human ORFX protein
720	6	0.7	69	22	AAU54247	Human bone marrow
721	6	0.7	69	22	AAU54247	Peptide #947 enco
722	6	0.7	69	22	AAU54247	Peptide #967 enco
723	6	0.7	69	22	AAU54247	Peptide #6269 enco
724	6	0.7	69	22	AAU54247	Peptide #921 enco
725	6	0.7	69	22	AAU54247	Human peptide enco
726	6	0.7	69	22	AAU54247	Human ORFX protein
727	6	0.7	69	22	AAU54247	Human bone marrow
728	6	0.7	69	22	AAU54247	Peptide #947 enco
729	6	0.7	69	22	AAU54247	Peptide #967 enco
730	6	0.7	69	22	AAU54247	Peptide #6269 enco
731	6	0.7	69	22	AAU54247	Peptide #921 enco
732	6	0.7	69	22	AAU54247	Human peptide enco
733	6	0.7	69	22	AAU54247	Human ORFX protein
734	6	0.7	69	22	AAU54247	Human bone marrow
735	6	0.7	69	22	AAU54247	Peptide #947 enco
736	6	0.7	69	22	AAU54247	Peptide #967 enco
737	6	0.7	69	22	AAU54247	Peptide #6269 enco
738	6	0.7	69	22	AAU54247	Peptide #921 enco
739	6	0.7	69	22	AAU54247	Human peptide enco
740	6	0.7	69	22	AAU54247	Human ORFX protein

814	6	0.7	87	21	AA612343	Zea mays protein f
815	6	0.7	87	21	AA633517	Arabidopsis thalia
816	6	0.7	87	22	AAU57419	Propionibacterium
817	6	0.7	87	22	ABG11467	Novel human diagno
818	6	0.7	87	22	ABG30017	Novel human diagno
819	6	0.7	88	22	AAU57235	Propionibacterium
820	6	0.7	88	22	ABG12946	Novel human diagno
821	6	0.7	88	22	AAAM5817	Human immune/haema
822	6	0.7	88	23	ABP35030	Human structural p
823	6	0.7	88	23	ABP35030	Human structural p
824	6	0.7	89	21	AAAB38255	Listeria monocytog
825	6	0.7	89	21	AAAB38257	Human secreted pro
826	6	0.7	89	22	ABG11062	Human secreted pro
827	6	0.7	89	22	ABP06314	Novel human diagno
828	6	0.7	90	19	AAAM80333	Human ORFX protein
829	6	0.7	90	21	AAAG08931	Oxidase amino acid
830	6	0.7	90	21	AAAG27162	Arabidopsis thalia
831	6	0.7	90	22	AAU56826	Zea mays protein f
832	6	0.7	90	22	ABP29079	Propionibacterium
833	6	0.7	90	22	ABP38241	Peptide #1730 enco
834	6	0.7	90	22	AAU20972	Peptide #5747 enco
835	6	0.7	90	22	AAAM58867	Human novel foetal
836	6	0.7	90	22	AAAM67421	Human brain expres
837	6	0.7	90	22	AAAM90449	Human bone marrow
838	6	0.7	90	22	AAAM15249	Human immune/haema
839	6	0.7	90	22	AAAM19033	Peptide #1683 enco
840	6	0.7	90	22	AAAM27710	Peptide #5467 enco
841	6	0.7	90	22	AAAM31674	Peptide #1747 enco
842	6	0.7	90	22	AAAM02993	Peptide #5711 enco
843	6	0.7	90	22	AAU20585	Peptide #1675 enco
844	6	0.7	90	22	ABP77825	Human secreted pro
845	6	0.7	90	23	ABG37044	Amino acid sequenc
846	6	0.7	91	19	AAAM60958	Human peptide enco
847	6	0.7	91	21	AAAG20176	Streptococcus pneu
848	6	0.7	91	21	AAAG33052	Arabidopsis thalia
849	6	0.7	91	21	AAAG33055	Zea mays protein f
850	6	0.7	91	21	AAAG33059	Zea mays protein f
851	6	0.7	91	22	ABP68643	Drosophila melanog
852	6	0.7	92	21	AAAB33250	Eucalyptus grandis
853	6	0.7	92	22	AAU60374	Propionibacterium
854	6	0.7	92	22	AAU31593	Novel human secret
855	6	0.7	92	23	ABP31138	Human synthase-11k
856	6	0.7	93	22	AAAM83247	Human immune/haema
857	6	0.7	93	22	AAAM41804	Human polypeptide
858	6	0.7	93	23	ABP00429	Human ORFX protein
859	6	0.7	93	23	ABP10159	Human ORFX protein
860	6	0.7	94	23	ABP80046	Mammary cancer 1 a
861	6	0.7	95	21	AAAG12453	Zea mays protein f
862	6	0.7	95	22	AAU67445	Propionibacterium
863	6	0.7	95	22	ABG09209	Novel human diagno
864	6	0.7	95	22	AAO02377	Human polypeptide
865	6	0.7	95	23	ABP34436	Human glycoprotein
866	6	0.7	96	21	AAAG40831	Zea mays protein f
867	6	0.7	96	22	AAO04390	Human polypeptide
868	6	0.7	96	23	ABP34208	Human ORFX181 prot
869	6	0.7	96	23	ABP09306	Human ORFX protein
870	6	0.7	97	21	AAAB45540	Human S100A10 prot
871	6	0.7	97	21	AAAY93605	Protein encoded by
872	6	0.7	97	22	AAU51033	Propionibacterium
873	6	0.7	97	22	ABP16109	Human nervous syst
874	6	0.7	97	22	AAAM50210	Human plasminogen
875	6	0.7	97	22	AAAB71655	Human colon associ
876	6	0.7	97	23	ABP25789	Streptococcus poly
877	6	0.7	98	18	AAAY11296	S. pneumoniae SPOU
878	6	0.7	98	21	AAAB25335	Plus radiata cell
879	6	0.7	98	21	AAAG40830	Zea mays protein f
880	6	0.7	98	21	AAAG00891	Human secreted pro
881	6	0.7	98	21	AAAY59245	Tiliapia IGF-II E p
882	6	0.7	98	22	AAU67006	Propionibacterium
883	6	0.7	98	22	AAU55100	Protein encoded by
884	6	0.7	98	22	AAAB84210	Human secreted pro
885	6	0.7	98	22	AAAG73408	Human gene 14-enco
886	6	0.7	98	23	ABG64233	Human albumin fusi
<hr/>						
887	6	0.7	98	23	ABP05302	Human ORFX protein
888	6	0.7	98	23	ABP06791	Human sailor trans
889	6	0.7	99	20	AAAY5094	Chlamydia pneumoni
890	6	0.7	99	22	AAAG1895	S. epidermidis ope
891	6	0.7	99	23	ABP07507	Human ORFX protein
892	6	0.7	100	21	AAAG49641	Arabidopsis thalia
893	6	0.7	100	21	AAAG61714	Arabidopsis thalia
894	6	0.7	100	22	AAU44981	Propionibacterium
895	6	0.7	100	22	ABG11918	Novel human diagno
896	6	0.7	101	20	AAAY34809	Chlamydia pneumoni
897	6	0.7	101	21	AAAG24035	Arabidopsis thalia
898	6	0.7	101	21	AAAG02231	Human secreted pro
899	6	0.7	101	22	ABP69053	Drosophila melanog
900	6	0.7	101	22	ABP35136	Peptide #2642 enco
901	6	0.7	101	22	ABP20558	Protein #2557 enco
902	6	0.7	101	22	AAAM68332	Human bone marrow
903	6	0.7	101	22	AAAM28640	Peptide #2677 enco
904	6	0.7	101	22	AAAM03878	Peptide #2560 enco
905	6	0.7	101	22	AAAG92336	C glutamicum prote
906	6	0.7	101	23	ABG37884	Human peptide enco
907	6	0.7	102	21	AAAG14367	Arabidopsis thalia
908	6	0.7	102	21	AAAG28001	Arabidopsis thalia
909	6	0.7	102	21	AAAG31521	Arabidopsis thalia
910	6	0.7	102	21	AAAG41892	Arabidopsis thalia
911	6	0.7	102	21	AAAG47020	Arabidopsis thalia
912	6	0.7	102	22	ABG17720	Novel human diagno
913	6	0.7	102	22	AAAM93300	Human polypeptide
914	6	0.7	102	22	AAO03306	Human polypeptide
915	6	0.7	102	22	AAO07741	Human polypeptide
916	6	0.7	103	21	AAAB40839	Human ORFX ORF603
917	6	0.7	103	21	AAAB41321	Human ORFX ORF1085
918	6	0.7	103	21	AAAB07688	A human interleuk1
919	6	0.7	103	21	AAAG27367	Arabidopsis thalia
920	6	0.7	103	21	AAAG49640	Arabidopsis thalia
921	6	0.7	103	22	AAAM86927	Human immune/haema
922	6	0.7	103	22	AAAB60226	Pseudomonas mendoc
923	6	0.7	103	23	ABP43031	Human ovarian anti
924	6	0.7	103	23	ABP02144	Human ORFX protein
925	6	0.7	103	23	ABP03730	Human ORFX protein
926	6	0.7	104	20	AAAY29093	Human mammary asso
927	6	0.7	104	21	AAAG40961	Zea mays protein f
928	6	0.7	104	22	AAAM96567	Human reproductive
929	6	0.7	104	22	AAO02347	Human polypeptide
930	6	0.7	104	23	ABP32685	Human ORF1658 prot
931	6	0.7	104	23	ABP05385	Human sailor trans
932	6	0.7	105	21	AAAB15396	E. coli metJ mutan
933	6	0.7	105	21	AAAG35923	Zea mays protein f
934	6	0.7	105	21	AAAG00194	Human secreted pro
935	6	0.7	105	22	AAU52438	Propionibacterium
936	6	0.7	105	22	AAO00368	Human polypeptide
937	6	0.7	105	23	AAO211778	Lung-specific amia
938	6	0.7	106	21	AAAB41441	Human ORFX ORF1205
939	6	0.7	106	21	AAAG33516	Arabidopsis thalia
940	6	0.7	106	22	AAU00893	Human cancer relat
941	6	0.7	106	23	ABP33606	Human ORF2579 prot
942	6	0.7	106	23	AAAB47869	Mature placenta-de
943	6	0.7	107	13	AAAR22852	Promoter/secretion
944	6	0.7	107	14	AAAR47658	Promoter-secretion
945	6	0.7	107	17	AAAR9708	PKTH1797-encoded p
946	6	0.7	107	21	AAAG18825	Zea mays protein f
947	6	0.7	107	21	AAAG58560	Arabidopsis thalia
948	6	0.7	107	22	AAU40317	Propionibacterium
949	6	0.7	107	22	ABP33413	Propionibacterium
950	6	0.7	107	22	ABP33413	Peptide #4064 enco
951	6	0.7	107	22	AAAB36627	Peptide #4133 enco
952	6	0.7	107	22	AAAM69781	Human brain expres
953	6	0.7	107	22	AAAM17606	Human bone marrow
954	6	0.7	107	22	AAAB35226	Peptide #4040 enco
955	6	0.7	107	23	ABG39413	Human peptide enco
956	6	0.7	107	23	AAAM48199	Human zinc finger
957	6	0.7	108	21	AAAG15029	Arabidopsis thalia
958	6	0.7	108	21	AAAG56859	Arabidopsis thalia
959	6	0.7	108	22	ABP69222	Drosophila melanog

960	6	0.7	108	22	AAU53021	Propionibacterium
961	6	0.7	108	22	AAOI0921	Human polypeptide
962	6	0.7	108	23	ABP04212	Human ORF protein
963	6	0.7	109	20	AAV13029	Human secreted pro
964	6	0.7	109	22	AAO01136	Human polypeptide
965	6	0.7	109	23	ABP03124	Human ORF protein
966	6	0.7	110	21	AAU51520	Arabidopsis thalia
967	6	0.7	110	21	AAU51520	Arabidopsis thalia
968	6	0.7	110	21	AAU51520	Arabidopsis thalia
969	6	0.7	110	21	AAU51520	Arabidopsis thalia
970	6	0.7	110	21	AAU51520	Arabidopsis thalia
971	6	0.7	110	21	AAU51520	Arabidopsis thalia
972	6	0.7	110	21	AAU51520	Arabidopsis thalia
973	6	0.7	110	21	AAU51520	Arabidopsis thalia
974	6	0.7	110	21	AAU51520	Arabidopsis thalia
975	6	0.7	110	21	AAU51520	Arabidopsis thalia
976	6	0.7	110	21	AAU51520	Arabidopsis thalia
977	6	0.7	110	21	AAU51520	Arabidopsis thalia
978	6	0.7	110	21	AAU51520	Arabidopsis thalia
979	6	0.7	110	21	AAU51520	Arabidopsis thalia
980	6	0.7	110	21	AAU51520	Arabidopsis thalia
981	6	0.7	110	21	AAU51520	Arabidopsis thalia
982	6	0.7	110	21	AAU51520	Arabidopsis thalia
983	6	0.7	110	21	AAU51520	Arabidopsis thalia
984	6	0.7	110	21	AAU51520	Arabidopsis thalia
985	6	0.7	110	21	AAU51520	Arabidopsis thalia
986	6	0.7	110	21	AAU51520	Arabidopsis thalia
987	6	0.7	110	21	AAU51520	Arabidopsis thalia
988	6	0.7	110	21	AAU51520	Arabidopsis thalia
989	6	0.7	110	21	AAU51520	Arabidopsis thalia
990	6	0.7	110	21	AAU51520	Arabidopsis thalia
991	6	0.7	110	21	AAU51520	Arabidopsis thalia
992	6	0.7	110	21	AAU51520	Arabidopsis thalia
993	6	0.7	110	21	AAU51520	Arabidopsis thalia
994	6	0.7	110	21	AAU51520	Arabidopsis thalia
995	6	0.7	110	21	AAU51520	Arabidopsis thalia
996	6	0.7	110	21	AAU51520	Arabidopsis thalia
997	6	0.7	110	21	AAU51520	Arabidopsis thalia
998	6	0.7	110	21	AAU51520	Arabidopsis thalia
999	6	0.7	110	21	AAU51520	Arabidopsis thalia
1000	6	0.7	110	21	AAU51520	Arabidopsis thalia

## ALIGNMENTS

## RESULT 1

AA95987 standard: Protein: 919 AA.

AA95987:

05-DEC-2000 (first entry)

Moraxella catarrhalis BASB081 protein.

BASB081: Infection; otitis media; pneumonia; sinusitis; inflammation; therapy; antibacterial; antiinflammatory; vaccine;

Moraxella catarrhalis.

Key Location/Qualifiers

Peptide 1..30 /label= Signal\_peptide

Protein 31..919 /label= Mature\_protein

Misc-difference 365 /note= "Ile in sequence of AA95988"

MO200052042-A1.

08-SEP-2000.

XX	23-FEB-2000: 2000MO-EP01468.
PF	26-FEB-1999: 99GB-0004559.
XX	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
PA	Ruelle J.
XX	WPI: 2000-587296/55.
DR	N-PSDB: AAA50536.
XX	New BASB081 polypeptides from Moraxella catarrhalis and polynucleotides
PT	encoding the polypeptides used for treating infections, or as a vaccine
PT	for preventing infections, especially those caused by M. catarrhalis
XX	Claim 1: Page 60; 97pp: English.
CC	The present sequence is that of the Moraxella catarrhalis strain
CC	Mc2391 (ATCC 43617) BASB081 protein, which is related by amino acid
CC	sequence homology to Neisseria meningitidis omp5 outer membrane
CC	protein. It shows 99.9% homology in the mature region to a
CC	provides (see AA95988) deduced from a PCR product. The invention
CC	host cells, and a process for producing a BASB081 polypeptide,
CC	polypeptide or polynucleotide, and optionally at least 1 other M.
CC	catarrhalis antigen. A method for diagnosing a M. catarrhalis
CC	infection involves identifying a BASB081 polypeptide, or an antibody
CC	that is immunospecific for it, in a sample. A therapeutic
CC	composition useful in treating M. catarrhalis diseases in humans
CC	disease can be a bacterial infection, e.g. otitis media in infants
CC	and children, pneumonia in elders, sinusitis, nosocomial
CC	infections and invasive diseases, chronic otitis media with hearing
CC	loss, fluid accumulation in the middle ear, auditive nerve damage,
CC	delayed speech learning, upper respiratory tract infection, and
CC	inflammation of the middle ear.
XX	Sequence 919 AA:
SQ	

Query Match	100.0%; Score 919; DB 21; Length 919;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 919; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MSKPVLFANRSEFMPVALAATPLMTSQALAOQNNPANIINHVAHDTATNOAKAGNPVL 60
DB	1 MSKPVLFANRSEFMPVALAATPLMTSQALAOQNNPANIINHVAHDTATNOAKAGNPVL 60
QY	61 LTPEDIOARLNAAGLNAPQSOALDVNFDOSPTSRIGESPPGLDMVSEETPTSL 120
DB	61 LTPEDIOARLNAAGLNAPQSOALDVNFDOSPTSRIGESPPGLDMVSEETPTSL 120
QY	121 EELFAQESTREKINPNDYIPYEQEQPNSEVVVPTLPPEKGLIKRLAFLNDGVNKKV 180
DB	121 EELFAQESTREKINPNDYIPYEQEQPNSEVVVPTLPPEKGLIKRLAFLNDGVNKKV 180
QY	181 PRLKAFYQSSQSGETSAIGSSHOKTEPYANIKALEDTIOESAMDNGSIPRLROFATV 240
DB	181 PRLKAFYQSSQSGETSAIGSSHOKTEPYANIKALEDTIOESAMDNGSIPRLROFATV 240
QY	241 AARAVGYDDIDLSIRNSIGEVYIHDLEPYIDYRAVEVGEADKAFITVADEVP 300
DB	241 AARAVGYDDIDLSIRNSIGEVYIHDLEPYIDYRAVEVGEADKAFITVADEVP 300
QY	301 LLIGDVFHGHGKETKKKLLENASAEHGYFDGRWLDRSVVILPDMADVSLTYDTGTQR 360
DB	301 LLIGDVFHGHGKETKKKLLENASAEHGYFDGRWLDRSVVILPDMADVSLTYDTGTQR 360
QY	361 FDEVEVFTTIDPKNTLTTPDKLPVRELLBOLLTYNMGEEANLQAVNALSNDLIATRYF 420
DB	361 FDEVEVFTTIDPKNTLTTPDKLPVRELLBOLLTYNMGEEANLQAVNALSNDLIATRYF 420

```
QY 421 NMVNTEIVPEREIONDOVSFEQSSSRTEPAQVDESTLEPIVETVELTDGILMDISPI 480
Db 421 NMVNTEIVPEREIONDOVSFEQSSSRTEPAQVDESTLEPIVETVELTDGILMDISPI 480
QY 481 EFSASNLIDDKLNLVAAKARHLTMDPDRVLAINHDDGVNRSLIGRISDAVSAVARAILP 540
Db 481 EFSASNLIDDKLNLVAAKARHLTMDPDRVLAINHDDGVNRSLIGRISDAVSAVARAILP 540
QY 541 DESENEVIDLPERTALANRRTPADVYQSKKVPPLYVFVASDKPRDGOIGLGWSDTGTRLY 600
Db 541 DESENEVIDLPERTALANRRTPADVYQSKKVPPLYVFVASDKPRDGOIGLGWSDTGTRLY 600
QY 601 TKFEHNLINRDGOQAQAEELRLSEDKKGVKLXAKRPLSHRLNDQRLATLGOQEVFGHSTN 660
Db 601 TKFEHNLINRDGOQAQAEELRLSEDKKGVKLXAKRPLSHRLNDQRLATLGOQEVFGHSTN 660
QY 661 GFDLSRTLEHEHSRSIIQNGWNRTYSRLRYRLDKLTKTQAPPETMODLPVDFVNGKPSQE 720
Db 661 GFDLSRTLEHEHSRSIIQNGWNRTYSRLRYRLDKLTKTQAPPETMODLPVDFVNGKPSQE 720
QY 721 ALLAGVAHVHTVADNLVNPARGYRQRYSLSEVSSGLVSDANMAIARAGISGVYSEGDNAY 780
Db 721 ALLAGVAHVHTVADNLVNPARGYRQRYSLSEVSSGLVSDANMAIARAGISGVYSEGDNAY 780
QY 781 GSNRAHQMTGGIOAGYIWSNENHVPYRLRFAGGDOSTINGYAHDSLSPISDKGYLTGCG 840
Db 781 GSNRAHQMTGGIOAGYIWSNENHVPYRLRFAGGDOSTINGYAHDSLSPISDKGYLTGCG 840
QY 841 VLAVGTAENVYEFMKDLRLAVFGDIGNAYDKGFTNDTKIGAGVGRNASVGOVRVAVAT 900
Db 841 VLAVGTAENVYEFMKDLRLAVFGDIGNAYDKGFTNDTKIGAGVGRNASVGOVRVAVAT 900
QY 901 GYKEEGNPILKHFICTPF 919
Db 901 GYKEEGNPILKHFICTPF 919

RESULT 2
AAY95988
ID AAY95988 standard; Protein; 889 AA.
XX
AC AAY95988;
XX
DT 05-DEC-2000 (first entry)
XX
DE Moraxella catarrhalis BASB081 mature protein.
XX
KW BASB081; infection; otitis media; pneumonia; sinusitis;
KW inflammation; therapy; antibacterial; antinflammatory; vaccine;
KW diagnosis.
XX
OS Moraxella catarrhalis.
XX
FH Key Location/Qualifiers
FT Misc-difference 335 /note="Val in sequence of AAY95987"
PI WO200052042-A1.
PN 08-SEP-2000.
PD 23-FEB-2000; 2000WO-EP01468.
PF 26-FEB-1999; 99GB-0004559.
PR (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA Ruelle J;
XX
XX MPI; 2000-587296/55.
XX DR N-PSDB; AAA50537.
XX PT New BASB081 polypeptides from Moraxella catarrhalis and polynucleotides
```

```
PT encoding the polypeptides used for treating infections, or as a vaccine
PT for preventing infections, especially those caused by M. catarrhalis -
XX
XX Claim 1; Page 61; 97pp; English.
XX
CC The present sequence is that of the Moraxella catarrhalis strain
CC Mc2391 (ATCC 43617) BASB081 mature protein. BASB081 is related by
CC amino acid sequence homology to Neisseria meningitidis omp5 outer
CC membrane protein. The sequence was deduced from PCR-amplified DNA
CC (see AAA50537), and shows 99.9% homology to the mature region of a
CC BASB081 gene translation product (see AAY95987). The invention
CC provides BASB081 polypeptides, polynucleotides, expression vectors,
CC host cells, and a process for producing a BASB081 polypeptide. Also
CC provided are vaccine compositions comprising a BASB081 polypeptide
CC or polynucleotide, and optionally at least 1 other M. catarrhalis
CC antigen. A method for diagnosing a M. catarrhalis infection
CC involves identifying a BASB081 polypeptide, or an antibody that is
CC immunospecific for it, in a sample. A therapeutic composition
CC useful in treating M. catarrhalis diseases in humans comprises an
CC antibody directed against a BASB081 polypeptide. The disease can
CC be a bacterial infection, e.g. otitis media in infants and children,
CC pneumonia in elderslies, sinusitis, nosocomial infections and
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed
CC speech learning, upper respiratory tract infection, and inflammation
CC of the middle ear.
XX
SQ Sequence 889 AA;
Query Match 85.7%; Score 788; DB 21; Length 889;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 888; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 31 QONNPANIIHVPAHDTAINDQAKAGNPVLLTPEQIARLNAAGINKPQSOALDVYVNF 90
Db 1 QONNPANIIHVPAHDTAINDQAKAGNPVLLTPEQIARLNAAGINKPQSOALDVYVNF 90
QY 91 DQSPISRIEGOSPPLGLIDMSYIEETTPLSLELFAQSTENGINPNDIYPYQGEQPNSE 150
Db 61 DQSPISRIEGOSPPLGLIDMSYIEETTPLSLELFAQSTENGINPNDIYPYQGEQPNSE 150
QY 151 VVVPPTLEPEKPGILKRLVYARLFPNDGVNKPRLAKFYQSSQSETSAIGSSHOKTEPER 210
Db 121 VVVPPTLEPEKPGILKRLVYARLFPNDGVNKPRLAKFYQSSQSETSAIGSSHOKTEPER 210
QY 211 NIKALEDITQESAMDINGSIPLRQATALVAARAVGYVDIDLSITRNSIGEVYIINDLG 270
Db 181 NIKALEDITQESAMDINGSIPLRQATALVAARAVGYVDIDLSITRNSIGEVYIINDLG 270
QY 271 EPVYIDYRAVEVRGEGADDAFTTVADEVPLLIDGVFHGKYEYKKNLIENASAEHGTFD 330
Db 241 EPVYIDYRAVEVRGEGADDAFTTVADEVPLLIDGVFHGKYEYKKNLIENASAEHGTFD 330
QY 331 GRWLDRSVDVILPNTADVSLIYDTGOYRDEVVFETTOPKTNOLTPDPKLLVKKRELL 390
Db 301 GRWLDRSVDVILPNTADVSLIYDTGOYRDEVVFETTOPKTNOLTPDPKLLVKKRELL 390
QY 391 EQLLTVMNGEAYNLQAAVRALSNDLIATRYFNMVTEIVFEREIONDOVSFEQSSSRT 450
Db 361 EQLLTVMNGEAYNLQAAVRALSNDLIATRYFNMVTEIVFEREIONDOVSFEQSSSRT 450
QY 451 EPAQVDESTLEPIVETVELTDGILMDISPIEFSASNLIDDKLNLVAAKARHLTMDPDRV 510
Db 421 EPAQVDESTLEPIVETVELTDGILMDISPIEFSASNLIDDKLNLVAAKARHLTMDPDRV 510
QY 511 LAIHHDDGVNRSLIGRISDAVSAVARAILPDESENEVIDLPERTALANRRTPADVYQSKK 570
Db 481 LAIHHDDGVNRSLIGRISDAVSAVARAILPDESENEVIDLPERTALANRRTPADVYQSKK 570
QY 571 VPLVVFVASDKPRDGOIGLGWSDTGTRLYTKFEHNLINRDGOQAQAEELRLSEDKKGVKL 630
Db 541 VPLVVFVASDKPRDGOIGLGWSDTGTRLYTKFEHNLINRDGOQAQAEELRLSEDKKGVKL 630
```

```

QY 631 YATKPLSHPLNDOLRATIGQOEYFGHSTNGFDLSTRLEHEISRSIIIONGMNRTYSR 690
DB 601 YATKPLSHPLNDOLRATIGQOEYFGHSTNGFDLSTRLEHEISRSIIIONGMNRTYSR 660
QY 691 YRLDKLKTQAPPTWODLPVDFVNGKPSOEALLAGVAHVKTADLVNPMRCYRQYSLE 750
DB 661 YRLDKLKTQAPPTWODLPVDFVNGKPSOEALLAGVAHVKTADLVNPMRCYRQYSLE 720
QY 751 VSSSGIVSANNALIRAGISGYISFGDNAYGSMRAHQMTGTGIOAGYIMSDNENHVPYRLR 810
DB 721 VSSSGIVSANNALIRAGISGYISFGDNAYGSMRAHQMTGTGIOAGYIMSDNENHVPYRLR 780
QY 811 FFAGGDOSIRGYAHDLSPISDKGYLTGGGYLAVGTAETAYNEFKKDLRLAVFGDIGNAYD 870
DB 781 FFAGGDOSIRGYAHDLSPISDKGYLTGGGYLAVGTAETAYNEFKKDLRLAVFGDIGNAYD 840
QY 871 KGTMDTKIGAGVGRASPVGOVRVDVANGVKEGPNIKLHPIGTF 919
DB 841 KGTMDTKIGAGVGRASPVGOVRVDVANGVKEGPNIKLHPIGTF 889

```

## RESULT 3

```

AA95821
ID AAY95821 standard; Protein: 576 AA.
AC
XX
XX

```

```

DT 07-NOV-2000 (first entry)
XX

```

```

DE Haemophilus influenza ncti strain 289 BASB067 protein.
XX

```

```

KW BASB067; Outer membrane protein; antigen; vaccine; antidiabetic;
antibacterial; screening; infection; diagnosis; therapy.
OS
XX

```

```

XX Haemophilus influenzae.
XX

```

```

FH Key
FT Location/Qualifiers
FT Peptide 1..20
FT Protein /label= Signal-Peptide
FT 21..576
FT /label= Mature-protein
XX

```

```

PN WO200047737-A1.
XX

```

```

PD 17-AUG-2000.
XX

```

```

PF 04-FEB-2000; 2000WO-EP00887.
XX

```

```

PR 09-FEB-1999; 99GB-0002880.
XX

```

```

PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX

```

```

PI Ruelle J, Thonnard J;
XX

```

```

DR WPI: 2000-515059/46.
XX

```

```

N-PSDB; AAA50270.
XX

```

```

PS BASB067 polypeptide and polynucleotide from Haemophilus influenzae are
used for diagnosing and treating H. influenzae infections.
XX
Claim 1; Page 81-82; 87pp; English.
XX

```

```

CC The present sequence is that of outer membrane protein BASB067 of
CC Haemophilus influenzae non-typable (Ncti) strain 289. BASB067 is
CC a surface expressed protein that is recognised by the immune system.
CC It shows homology to the protective surface antigen D15 of
CC H. influenzae and has a similar secondary structure. Its
CC N-terminal domain is predicted to contain a mixture of alpha-helix
CC and beta-strands, and could be used as a vaccine antigen. The
CC C-terminal domain is predicted to form a beta-barrel composed of
CC anti-parallel, amphipathic beta-strands. The external loops of
CC the beta-barrel of integral outer membrane proteins frequently
CC contain immunodominant B-cell epitopes, making the C-terminal

```

```

CC domain of BASB067 a strong candidate vaccine antigen. The
CC invention relates to recombinant materials and methods for the
CC production of BASB067 polypeptides and polynucleotides, for use
CC especially in therapeutic and prophylactic vaccines. It also
CC relates to methods for using such polypeptides and polynucleotides
CC in the prevention and treatment of microbial diseases, in diagnostic
CC assays for detecting diseases associated with microbial infections,
CC and assays for detecting expression or activity of BASB067
CC polypeptides or polynucleotides. Antibodies raised against
CC the mature portion of this BASB067 polypeptide can be used to treat
CC humans with H. influenzae disease.
XX
SQ Sequence 576 AA:

```

```

Query Match 1.2%; Score 11; DB 21; Length 576;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 882 GGVGRMASPVG 892
DB 537 GGVGRMASPVG 547

```

## RESULT 4

```

AA95820
ID AAY95820 standard; Protein: 578 AA.
AC
XX
XX

```

```

AC AAY95820;
XX

```

```

DT 07-NOV-2000 (first entry)
XX

```

```

DE Haemophilus influenza strain Rd KW20 BASB067 protein.
XX

```

```

KW BASB067; Outer membrane protein; antigen; vaccine; antidiabetic;
antibacterial; screening; infection; diagnosis; therapy.
XX

```

```

XX Haemophilus influenzae.
XX

```

```

FH Key
FT Location/Qualifiers
FT Peptide 1..22
FT Protein /label= Signal-Peptide
FT 23..578
FT Domain /label= Mature-protein
FT 23..236
FT /note= "N-terminal domain"
FT 237..578
FT /note= "C-terminal domain"
XX

```

```

PN WO200047737-A1.
XX

```

```

PD 17-AUG-2000.
XX

```

```

PF 04-FEB-2000; 2000WO-EP00887.
XX

```

```

PR 09-FEB-1999; 99GB-0002880.
XX

```

```

PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX

```

```

PI Ruelle J, Thonnard J;
XX

```

```

DR WPI: 2000-515059/46.
XX

```

```

N-PSDB; AAA50269.
XX

```

```

PS BASB067 polypeptide and polynucleotide from Haemophilus influenzae are
used for diagnosing and treating H. influenzae infections.
XX
Claim 1; Page 78-80; 87pp; English.
XX

```

```

CC The present sequence is that of outer membrane protein BASB067 of
CC Haemophilus influenzae strain Rd KW20. BASB067 is a surface
CC expressed protein that is recognised by the immune system. It
CC shows 23% identity to the protective surface antigen D15 of
CC H. influenzae and has a similar secondary structure. Its

```

CC N-terminal domain is predicted to contain a mixture of alpha-helix  
CC and beta-strands, and could be used as a vaccine antigen. The  
CC C-terminal domain is predicted to form a beta-barrel composed of  
CC anti-parallel, amphipathic beta-strands. The external loops of  
CC the beta-barrels of integral outer membrane proteins frequently  
CC contain immunodominant B cell epitopes, making the C-terminal  
CC domain of BASB067 a strong candidate vaccine antigen. The  
CC invention relates to recombinant materials and methods for the  
CC production of BASB067 polypeptides and polynucleotides, for use  
CC especially in therapeutic and prophylactic vaccines. It also  
CC relates to methods for using such polypeptides and polynucleotides  
CC in the prevention and treatment of microbial diseases, in diagnostic  
CC assays for detecting diseases associated with microbial infections,  
CC and assays for detecting expression or activity of BASB067  
CC polypeptides or polynucleotides. Antibodies raised against  
CC BASB067 can be used to treat humans with H. influenzae disease.  
XX  
SQ Sequence 578 AA;

Query Match 1.2%; Score 11; DB 21; Length 578;  
Best Local Similarity 100.0%; Pred. No. 0.071;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 882 GCGVRASPYG 892  
Db 539 GCGVRASPYG 549

## RESULT 5

AAY70579  
ID AAY70579 standard; Protein; 262 AA.

AC AAY70579;

DT 04-JUL-2000 (first entry)

DE Salmonella Pathogenicity Island 2 (SPI2) SseF protein.

XX  
XX Salmonella Pathogenicity Island 2 locus; SPI2 locus; antigen carrier;  
KW attenuation; gram-negative cell; vaccine; cytosolic; virulence; tumour;  
KW anti-arteriosclerotic; anti-Alzheimer's; bactericide; hepatotropic;  
KW anti-inflammatory; microbial infection; therapeutic; Salmonella infection;  
KW Helicobacter pylori; stomach cancer; Herpes virus; Chlamydia pneumoniae;  
KW Alzheimer's disease; arteriosclerosis; viral pathogen; Hepatitis virus;  
KW cervical cancer; sseF; effector; type III secretion system.  
XX  
XX

OS Salmonella typhimurium.

XX  
XX WO200014240-A2.

PD 16-MAR-2000.

XX  
XX 03-SEP-1999; 99WO-EP06514.

XX  
XX 04-SEP-1998; 98EP-0116827.

XX  
XX (CREA-) CREATOGEN BIOSCIENCES GMBH.

XX  
XX Hensel M, Guzman CA, Medina E, Apfel H, Hueck C;

XX  
XX WPI: 2000-256988/22.

DR N-PSDB; AAQ49164.

XX  
XX Attenuated gram-negative Salmonella cells, comprising inactivated genes  
PT in the SPI2 locus and useful for vaccinating against a range of  
PT disorders associated with microbial infections such as stomach and  
PT cervical cancers -  
XX

XX  
XX Claim 18; Fig 23F; 180pp; English.

XX  
XX The patent discloses attenuated gram-negative cells, especially  
CC Salmonella, in which at least 1 gene in the Salmonella Pathogenicity  
CC Island 2 (SPI2) locus has been inactivated resulting in attenuation/  
CC

CC reduction of virulence compared to the wild type cell. The attenuated  
CC cells are used as carriers for presenting bacterial, viral or tumour  
CC antigens to a host and are capable of expressing the antigen in a target  
CC cell, especially a macrophage. The cells may therefore be used for the  
CC preparation of a prophylactic or therapeutic composition for the  
CC treatment of a chronic disease caused by a bacterium or virus, e.g.  
CC Salmonella infection or a tumour. The cells may also be used to vaccinate  
CC against a range of bacterial and viral pathogens e.g. Helicobacter pylori  
CC (associated with stomach cancer), Chlamydia pneumoniae (associated with  
CC arteriosclerosis and Alzheimer's disease), Borrelia burgdorferi,  
CC Nanobacteria (found in the chronically diseased kidneys of patients  
CC with crystalline deposits), Hepatitis virus (causative agent of  
CC Hepatitis B and C and associated with liver cancer), Human papilloma  
CC virus (HPV) (associated with cervical cancer) or Herpes virus.  
CC The present sequence is the SseF protein, an effector of type III  
CC secretion system, from the SPI2 locus of Salmonella. Inactivation of the  
CC sse gene is useful for producing the attenuated cells.  
XX

Query Match 1.0%; Score 9; DB 21; Length 262;  
Best Local Similarity 100.0%; Pred. No. 3.7;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 PVLTPEOI 66  
Db 43 PVLTPEOI 51

## RESULT 6

AA41333  
ID AA41333 standard; Protein; 851 AA.

XX  
XX AA41333;

DT 22-APR-1994 (first entry)

DE 113 KD ISGF-3alpha.

XX  
XX 113 KD; 91 KD; 84 KD; ISGF-3alpha; interferon-related; receptor;  
KW recognition factor; gene family; translation protein; tyrosine;  
KW DNA binding protein; interferon-gamma; hairy cell leukaemia;  
KW interferon therapy; chronic viral hepatitis; phosphorylation;  
KW adjuvant therapy; tyrosine kinase.  
XX  
XX

OS Homo sapiens.

XX  
XX WO9319179-A.

XX  
XX 30-SEP-1993.

XX  
XX 19-MAR-1993; 93WO-US02569.

XX  
XX 19-MAR-1992; 92US-0854296.

XX  
XX 23-NOV-1992; 92US-0980498.

XX  
XX (UYRQ) UNIV ROCKEFELLER.

XX  
XX Darnell JE, Fu X, Schindler CW, Shuai K;

XX  
XX WPI: 1993-320745/40.

DR N-PSDB; AAQ49164.

XX  
XX Interferon receptor recognition factors - useful e.g. to treat  
PT viral hepatitis, hairy cell leukaemia and to potentiate interferon  
PT effects  
XX

XX  
XX Claim 17; Fig 1; 131pp; English.

XX  
XX The sequences given in AA41333-35 represent the 113 KD, 91 KD and 84 KD  
CC ISGF-3alpha proteins respectively. ISGF-3alpha is an interferon-  
CC related receptor recognition factor which comprises several  
CC substituents. The 113 KD, and the 91 and 84 KD proteins are derived  
CC

CC from two different but related genes. It is clear that a gene  
 CC family exists and further members are likely to be found. The 91 kD  
 CC protein has the capability of acting as a translation protein and as  
 CC a DNA binding protein in response to interferon-gamma stimulation.  
 CC These proteins participate in rapid phosphorylation and dephosphory-  
 CC lation during the course of, and as part of their activity. This  
 CC phosphorylation takes place in an interferon-dependant manner on  
 CC specified tyrosine residues. These proteins may be used in  
 CC conjunction with interferon therapy eg. to treat chronic viral  
 CC hepatitis, hairy cell leukemia and for use with interferon in  
 CC adjuvant therapy.

SO Sequence 851 AA:

Query Match 1.0%; Score 9; DB 14; Length 851;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ESTLEPVIE 465  
 |||||||||  
 DB 756 ESTLEPVIE 764

#### RESULT 7

ID AAR72077 standard; Protein; 851 AA.

AC AAR72077;

DT 27-SEP-1995 (first entry)

DE Recognition factor Stat13.

KW Signal transducer and activator of transcription; ISGF-3; STAT;

KM Stat13; receptor recognition factor; transcription factor;

KW cellular debilitation; derangement; dysfunction; interferon-alpha;

OS Homo sapiens.

PN W09508629-A.

PD 30-MAR-1995.

PE 26-SEP-1994; 94WO-US10849.

PR 24-SEP-1993; 93US-0126588.

PR 24-SEP-1993; 93US-0126595.

PR 11-MAR-1994; 94US-0212184.

PR 11-MAR-1994; 94US-0212185.

PA (UYRQ ) UNIV ROCKEFELLER.

PI Darnell JE, Schindler CW, Shuai K, Wen Z, Zhong Z;

DR WPI; 1995-139598/18.

DR N-PSDB; AA089335.

PT Receptor recognition factor implicated in transcriptional

PT stimulation of genes useful in drug screening assays and/or

PT for treating cellular debilitations, derangements and/or

PT dysfunctions, etc.

XX The sequences of cDNA encoding receptor recognition factors having

XX mol. wt. of 113 kDa (Stat13), 91 kDa (Stat92) and 84 kDa (Stat84)

XX are given in AA089335-37 and the deduced amino acid sequences of the

XX STAT proteins in AAR72077-79. These ISGF-3-derived proteins are

XX activated by binding of interferon-alpha (all 3 Stat proteins) or

XX interferon-gamma (Stat91) to cell receptors.

Query Match 1.0%; Score 9; DB 16; Length 851;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ESTLEPVIE 465  
 |||||||||  
 DB 756 ESTLEPVIE 764

#### RESULT 8

ID AAM03166 standard; Protein; 851 AA.

AC AAM03166;

DT 24-OCT-1996 (first entry)

DE Human STAT2.

KW STAT; STAT2; signal transducer and activator of transcription;

KW DNA binding protein; ligand; receptor; oncogenesis; inflammation;

KW autoimmune disease; antagonist; therapy; STAT13.

OS Homo sapiens.

PN Key Location/Qualifiers

FT Domain 396..506

FT /label= "DNA binding domain

FT /note= "Claim 3, page 11"

PN W09620954-A2.

PD 11-JUL-1996.

PF 28-DEC-1995; 95WO-US17025.

PR 06-JAN-1995; 95US-0369796.

PA (UYRQ ) UNIV ROCKEFELLER.

PI Darnell JE, Horvath CM, Wen Z, Zhong Z;

DR WPI; 1996-33941/33.

DR N-PSDB; AAT31275.

PT New STAT protein DNA-binding domain peptide(s) - useful for

PT diagnosing, preventing or treating cellular dysfunction, e.g.

PT oncogenesis; inflammation; parasitic disease or autoimmunity

XX Disclosure: Page 63-66; 138pp; English.

XX Signal transducer and activator of transcription (STAT) protein STAT2

XX (AAM03166), also known as STAT113, is a 113 kDa protein having a dual

XX purpose, i.e. signal transduction from ligand-activated receptor

XX kinase complexes followed by nuclear translocation and DNA binding

XX to activate transcription. The amino acid sequence of STAT2 was

XX deduced from a cDNA clone (AAT31275) derived from HeLa cells. STAT2

XX includes a DNA-binding domain (see also AAM03175) capable of both

XX receptor recognition and message delivery via DNA binding in a

XX receptor-ligand specific manner. STAT proteins and their DNA

XX binding domains (see also AAM03165, AAM03167-76) are useful for

XX screening antagonists used to inhibit STAT-mediated signal

XX transduction and activation of transcription.

SO Sequence 851 AA:

Query Match 1.0%; Score 9; DB 17; Length 851;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ESTLEPVIE 465  
 |||||||||



DB 756 ESTLEPVIE 764

RESULT 9  
AAB12374

ID AAB12374 standard; peptide; 851 AA.

XX AAB12374;

DT 08-NOV-2000 (first entry)

XX N-terminal domain of human STAT-2 protein.

DE STAT; signal transducer and activator of transcription; crystal;  
KW drug design; human.

XX Homo sapiens.

OS Key Location/Qualifiers

FT Region 4..9 /label= "Alpha helix 1"

FT Region 12..21 /label= "Alpha helix 2"

FT Region 19..21 /label= "3(10) helix of alpha helix 2"

FT Region 29..34 /label= "Alpha helix 3"

FT Region 36..41 /label= "Alpha helix 4"

FT Region 44..48 /label= "Alpha helix 5"

FT Region 53..76 /label= "Alpha helix 6"

FT Region 81..99 /label= "Alpha helix 7"

FT Region 102..122 /label= "Alpha helix 8"

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465

DB 756 ESTLEPVIE 764

RESULT 10

AAB19963

ID AAB19963 standard; Protein; 851 AA.

XX AAB19963;

DT 28-MAR-2001 (first entry)

XX Human signal transducer and activator of transcription STAT-2.

DE STAT-2; signal transducer and activator of transcription 2;

XX human; crystal; drug screening.

XX Homo sapiens.

OS Key Location/Qualifiers

FT Domain 139..315 /note= "coiled-coil domain"

FT Domain 316..485 /note= "DNA-binding domain"

FT Domain 486..575 /note= "linker domain"

FT Domain 576..680 /note= "SH2 domain"

FT Region 680..702 /note= "C-terminal tail segment"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

FT Modified-site 690 /note= "O-phosphorylated"

XX Sequence 851 AA;  
SQ  
Query Match 1.0%; Score 9; DB 22; Length 851;  
Best Local Similarity 100.0%; Pred No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 457 ESTLEPVIE 465  
DB 756 ESTLEPVIE 764  
IIIIIIII  
RESULT 11  
AAE14649  
ID AAE14649 standard; Protein; 851 AA.  
XX  
AC AAE14649;  
XX  
DT 16-JUL-2002 (first entry)  
XX  
DE Human STAT2 protein.  
XX  
KW Signal transducer and activator of transcription; STAT2;  
KW drug development; drug discovery; crystal; inflammation; allergy;  
KW asthma; leukaemia; anaemia; neutropenia; thrombocytopaenia;  
KW cancer; obesity; viral disease; growth retardation; human.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Domain 1..130  
FT /note= "Conserved N-terminal domain of the STAT family"  
XX  
PN US6312887-B1.  
XX  
PD 06-NOV-2001.  
XX  
PE 24-APR-2000; 2000US-0556273.  
XX  
PR 23-JAN-1998; 98US-0012710.  
XX  
PA (UVRQ ) UNIV ROCKEFELLER.  
XX  
PI Vinkemeier U, Moarefi I, Darnell JE, Kuriyan J;  
XX  
DR WPI: 2002-033337/04.  
XX  
PT Identifying compounds that bind to signal transducer and activator of  
PT transcription proteins, useful for the production of new drugs  
PS Example; Column 35-40; 44pp; English.  
XX  
XX The invention relates to methods for detecting compounds that bind to  
CC signal transducer and activator of transcription (STAT) proteins for  
CC the discovery and development of new drug compounds based on the  
CC structural properties of the protein crystal. The methods include:  
CC identifying a compound that binds to the N-terminal domain of a  
CC STAT protein, identifying a compound that enhances or diminishes the  
CC binding of the dimeric STAT proteins to each other and/or their nucleic  
CC acid binding site; or identifying a compound that enhances or diminishes  
CC the ability of STAT protein dimers to induce the expression  
CC of a gene operably under the control of a promoter containing at least  
CC two adjacent weak binding sites for STAT protein dimers. The methods  
CC are used for identifying new drugs. An antagonist of STAT N-terminal  
CC dimeric interactions that inhibits the binding of the STAT dimers to  
CC adjacent weak binding sites on a promoter of a gene, could be useful  
CC as drugs in the treatment of diseases, e.g. inflammation, allergy,  
CC asthma and leukaemia. On the other hand, an agonist of N-terminal  
CC dimeric interactions between STAT dimers, can be used as drugs in the  
CC treatment of diseases e.g. anaemia, neutropenia, thrombocytopaenia,  
CC cancer, obesity, viral diseases and growth retardation.  
XX The present sequence is human STAT2 protein.

SQ Sequence 851 AA;  
SQ  
Query Match 1.0%; Score 9; DB 23; Length 851;  
Best Local Similarity 100.0%; Pred No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 457 ESTLEPVIE 465  
DB 756 ESTLEPVIE 764  
IIIIIIII  
RESULT 12  
AAE15173  
ID AAE15173 standard; Protein; 851 AA.  
XX  
AC AAE15173;  
XX  
DT 07-MAR-2002 (first entry)  
XX  
DE Human Stat2 (signal transducer and activator of transcription) protein.  
XX  
KW Human; SOCS; suppressor of cytokine signalling; autoimmune disorder; Jak;  
KW Janus kinase; signal transducer and activator of transcription; Stat;  
KW transplant acceptance; systemic lupus erythematosus; glomerulonephritis;  
KW rheumatoid arthritis; Wegener's granulomatosis; chronic active hepatitis;  
KW atopic dermatitis; multiple sclerosis; myasthenia gravis; vasculitis;  
KW Crohn's disease; haemolytic anaemia; nephrotic syndrome; dermatological;  
KW diabetes mellitus; thyroiditis; inflammatory bowel disease; nephrotropic;  
KW immunosuppressive; antiinflammatory; antirheumatic; antiarthritic;  
KW hepatotropic.  
XX  
OS Homo sapiens.  
XX  
FH WO200179555-A2.  
FT 25-OCT-2001.  
XX  
PD 13-APR-2001; 2001WO-US12131.  
XX  
PR 14-APR-2000; 2000US-0549654.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Hancock WW, Ozkaynak E;  
XX  
DR WPI: 2002-034368/04.  
XX  
PT N-PSDB; AAD24305.  
XX  
PT Monitoring transplant acceptance or autoimmune disease, useful e.g. for  
PT assessing therapy, comprises measuring levels of Stat or their  
PT inhibitors  
PS Example; Fig 2; 218pp; English.  
XX  
XX The present invention relates to a method for monitoring acceptance of a  
CC transplant or an autoimmune disease in a mammal. The method comprising  
CC determining the amount of at least one of Stat4 (signal transducer and  
CC activator of transcription), Stat6, SOCS1 (suppressor of cytokine  
CC signalling) or SOCS3 mRNA or protein in a sample of the transplant, taken  
CC from the host or an affected tissue sample. Stats are activated by  
CC receptor-associated Janus kinases (Jaks) which include Jak1, Jak2, Tyk2,  
CC Jak3. The method is used to determine whether acceptance of a transplant  
CC has been induced or to determine if autoimmune disorders (systemic  
CC lupus erythematosus, glomerulonephritis, rheumatoid arthritis, Wegener's  
CC granulomatosis, chronic active hepatitis, atopic dermatitis, multiple  
CC sclerosis, myasthenia gravis, haemolytic anaemia, nephrotic syndrome,  
CC thyroiditis, diabetes mellitus, Crohn's disease, inflammatory bowel  
CC disease and vasculitis) are being treated successfully and may be used  
CC to adjust treatment regimes. The present sequence is human Stat2 protein.  
XX  
SQ Sequence 851 AA;  
SQ  
Query Match 1.0%; Score 9; DB 23; Length 851;

Best Local Similarity 100.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465  
|||||||  
Db 756 ESTLEPVIE 764

## RESULT 13

ABG25555  
ID ABG25555 standard; Protein: 114 AA.

AC ABG25555;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #25546.

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN M0200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB: AAS89742.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

PS Claim 20; SEQ ID NO 55914; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 114 AA;

Query Match 0.9%; Score 8; DB 22; Length 114;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 809 LRFAGCD 816

|||||||  
Db 52 LRFAGCD 59

## RESULT 14

ABB47876  
ID ABB47876 standard; Protein: 331 AA.

AC ABB47876;

DT 05-FEB-2002 (first entry)

DE Listeria monocytogenes protein #580.

KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KM vitamin B12; bacterial infection; disease.

OS Listeria monocytogenes.

PN M0200177335-A2.

PD 18-OCT-2001.

PF 11-APR-2001; 2001WO-FR01118.

PR 11-APR-2000; 2000FR-0004629.

PA (INSP ) INST PASTEUR.

PI Buchrieser C, Frangoul L, Couve E, Rusnlok C, Fsihi H, Dehoux P;  
PI Dussurget O, Chetoui F, Medjari H, Glaser P, Kunst F, Cossart P;  
PI Daniels J, Goebel W, Kreft J, Kunz M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Meduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voss H;

DR WPI: 2002-010914/01.

PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and  
PT related polypeptides -

PS Claim 6; SEQ ID NO 581; 192pp; French.

XX The present invention relates to the genome sequence of Listeria  
XX monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
XX it are useful for selecting probes and primers for detecting genes in L.  
XX monocytogenes and related organisms, and for studying genetic  
XX polymorphisms and other genomes. The present sequence is a protein  
XX encoded by the genome sequence of the present invention. Proteins  
XX expressed from the genome sequence are useful for raising specific  
XX antibodies, identification of L. monocytogenes and related organisms,  
XX for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
XX B12. The genome sequence and proteins encoded by it are also useful for  
XX selecting compounds that regulate gene expression and cell replication  
XX and modulate L. monocytogenes-related diseases. In addition, the genome  
XX sequence and proteins encoded by it are useful in pharmaceutical and  
XX vaccine compositions for the treatment or prevention of infections by L.  
XX monocytogenes and related organisms.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 331 AA;

Query Match 0.9%; Score 8; DB 23; Length 331;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 839 GOVLAVGT 846  
 Db 38 GOVLAVGT 45

## RESULT 15

ABG25558  
 ID ABG25558 standard; Protein; 371 AA.  
 AC  
 XX ABG25558;  
 XX

DT 18-FEB-2002 (first entry)  
 XX

DE Novel human diagnostic protein #25549.  
 XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.

PN WO200175067-A2.  
 XX

PD 11-OCT-2001.  
 XX

PF 30-MAR-2001: 2001WO-US08631.  
 XX

PR 31-MAR-2000: 2000US-0540217.  
 XX 23-AUG-2000: 2000US-0649167.  
 XX

PA (HYSE-) HYSEQ INC.  
 XX

PI Drmanac RT, Liu C, Tang YT;  
 XX

DR WPI: 2001-639362/73.  
 DR N-PSDB: AAS89745.  
 XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX

PS Claim 20; SEQ ID NO 55917; 103pp; English.  
 XX

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations in  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 371 AA;

Query Match 0.9%; Score 8; DB 22; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 809 LRFAGCD 816  
 Db 112 LRFAGCD 119

Db 185 LRFAGCD 192

## RESULT 16

ABG17710  
 ID ABG17710 standard; Protein; 474 AA.  
 AC  
 XX ABG17710;  
 XX

DT 18-FEB-2002 (first entry)  
 XX

DE Novel human diagnostic protein #17701.  
 XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.

PN WO200175067-A2.  
 XX

PD 11-OCT-2001.  
 XX

PF 30-MAR-2001: 2001WO-US08631.  
 XX

PR 31-MAR-2000: 2000US-0540217.  
 XX 23-AUG-2000: 2000US-0649167.  
 XX

PA (HYSE-) HYSEQ INC.  
 XX

PI Drmanac RT, Liu C, Tang YT;  
 XX

DR WPI: 2001-639362/73.  
 DR N-PSDB: AAS81897.  
 XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX

PS Claim 20; SEQ ID NO 48069; 103pp; English.  
 XX

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations in  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 474 AA;

Query Match 0.9%; Score 8; DB 22; Length 474;  
 Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 809 LRFAGCD 816  
 Db 112 LRFAGCD 119

```

RESULT 17
AAW56318
ID AAW56318 standard; Protein; 582 AA.
XX
AC AAW56318;
XX
DT 27-JUL-1998 (first entry)
XX
DE Infectious laryngotracheitis virus antigenic protein.
XX
KM Antigenic protein; Infectious laryngotracheitis virus; ILTV;
KM recombinant virus; fowlpox; avipox; chickenpox; vaccine; chicken;
KM pheasant; turkey.
XX
OS Infectious laryngotracheitis virus.
XX
PN WO9807866-A1.
XX
PD 26-FEB-1998.
XX
PF 21-AUG-1997; 97WO-JP02912.
XX
PR 21-AUG-1996; 96JP-0238580.
XX
PA (JAPG ) NIPPON ZEON KK.
XX
PI Okuda T, Tsuzaki Y;
XX
DR WPI: 1998-169170/15.
DR N-PSDB: AAW22805.
XX
PT Antigenic protein from Infectious laryngotracheitis virus and DNA
PT coding for it - which is incorporated into recombinant avian virus
PT for vaccine production
XX
PS Claim 1; Pages 38-41; 69pp; Japanese.
XX
CC The present sequence represents an antigenic protein originating in
CC infectious laryngotracheitis virus (ILTV). The DNA sequence can be
CC incorporated into recombinant viruses (such as fowlpox, avipox or
CC chickenpox virus) and used to produce vaccines. The vaccines are an
CC effective means of protecting birds such as chickens, pheasants or
CC turkeys against infectious laryngotracheitis.
XX
SQ Sequence 582 AA;

Query Match 0.9%; Score 8; DB 19; Length 582;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 213 KAALEDIT 220
Db 277 KAALEDIT 284

RESULT 18
ABB62368
ID ABB62368 standard; Protein; 1373 AA.
XX
AC ABB62368;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13896.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX

```

```

PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
DR N-PSDB: ABL06471.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure: SEQ ID NO 13896; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1373 AA;

Query Match 0.9%; Score 8; DB 22; Length 1373;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LTPEDIOA 68
Db 913 LTPEDIOA 920

RESULT 19
AAR44425
ID AAR44425 standard; Protein; 29 AA.
XX
AC AAR44425;
XX
DT 26-MAY-1994 (first entry)
XX
DE Mutant alpha-amylase signal peptide.
XX
KM Bacillus licheniformis; alpha-amylase; signal peptide;
KM peptidase; mature protein; recognition site; mutant;
KM cleavage site; interleukin; IL-3.
XX
OS Bacillus licheniformis.
XX
PN EP572088-A.
XX
PD 01-DEC-1993.
XX
PF 25-MAY-1993; 93EP-0201500.
XX
PR 25-MAY-1992; 92EP-0201492.
XX
PA (KONN ) GIST-BROCADES NV.
XX
PI Bonekamp AJ, Van Tilborg MWEW;
XX
DR WPI: 1993-379003/48.
XX

```

PT New mutant signal sequences with fewer potential cleavage sites -  
PT reduce processing ambiguity of signal peptidase(s)  
PS Claim 7; Page 15; 23pp; English.  
XX  
XX  
CC The two cleavage sites of B. licheniformis alpha-amylase signal  
CC peptide may give rise to two different mature proteins (AAR44429).  
CC To obtain a proper mature prod., the second recognition site  
CC has to be destroyed without interfering with the first recognition  
CC site. The combination of the B. licheniformis alpha-amylase  
CC signal sequence with human IL-3 shows a third possible cleavage  
CC site since mature IL-3 starts with Ala (AAR51699). In practice  
CC the third site is not used, as the signal peptidase is not  
CC active when the Ala-X-Ala is upstream of a pro.  
CC Mutant signal sequences with fewer potential cleavage sites  
CC reduce the processing ambiguity of signal peptidases.  
CC Mutated Bacillus alpha-amylase sequences in combination with  
CC IL-3 are illustrated in AAR44425-28.  
CC The oligonucleotide probes used to obtain the desired mutations  
CC are given in AAO52516-25.  
XX  
SQ Sequence 29 AA:  
  
Query Match 0.88; Score 7; DB 14; Length 29;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 166 KRLYARL 172  
|||||||  
Db 5 KRLYARL 11  
  
RESULT 20  
AAR44426  
ID AAR44426 standard; Protein: 29 AA.  
XX  
AC AAR44426;  
XX  
DT 26-MAY-1994 (first entry)  
XX  
DE Mutant alpha-amylase signal peptide.  
XX  
KW Bacillus licheniformis; alpha-amylase; signal peptide;  
KW peptidase; mature protein; recognition site; mutant;  
KW cleavage site; interleukin; IL-3.  
XX  
OS Bacillus licheniformis.  
XX  
PN EP572088-A.  
PD 01-DEC-1993.  
XX  
PF 25-MAY-1993; 93EP-0201500.  
XX  
PR 25-MAY-1992; 92EP-0201492.  
XX  
PA (KONN ) GIST-BROCADES NV.  
PI Bonekamp AJ, Van Tilborg MWEM;  
XX  
DR WPT; 1993-379003/48.  
XX  
PT New mutant signal sequences with fewer potential cleavage sites -  
PT reduce processing ambiguity of signal peptidase(s)  
PS Claim 7; Page 15; 23pp; English.  
XX  
XX The two cleavage sites of B. licheniformis alpha-amylase signal  
CC peptide may give rise to two different mature proteins (AAR44429).  
CC To obtain a proper mature prod., the second recognition site  
CC has to be destroyed without interfering with the first recognition  
CC site. The combination of the B. licheniformis alpha-amylase  
CC signal sequence with human IL-3 shows a third possible cleavage  
CC site. The combination of the B. licheniformis alpha-amylase  
CC signal sequence with human IL-3 shows a third possible cleavage

CC site since mature IL-3 starts with Ala (AAR51699). In practice  
CC the third site is not used, as the signal peptidase is not  
CC active when the Ala-X-Ala is upstream of a pro.  
CC Mutant signal sequences with fewer potential cleavage sites  
CC reduce the processing ambiguity of signal peptidases.  
CC Mutated Bacillus alpha-amylase sequences in combination with  
CC IL-3 are illustrated in AAR44425-28.  
CC The oligonucleotide probes used to obtain the desired mutations  
CC are given in AAO52516-25.  
XX  
SQ Sequence 29 AA:  
  
Query Match 0.88; Score 7; DB 14; Length 29;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 166 KRLYARL 172  
|||||||  
Db 5 KRLYARL 11  
  
RESULT 21  
AAR44427  
ID AAR44427 standard; Protein: 29 AA.  
XX  
AC AAR44427;  
XX  
DT 26-MAY-1994 (first entry)  
XX  
DE Mutant alpha-amylase signal peptide.  
XX  
KW Bacillus licheniformis; alpha-amylase; signal peptide;  
KW peptidase; mature protein; recognition site; mutant;  
KW cleavage site; interleukin; IL-3.  
XX  
OS Bacillus licheniformis.  
XX  
PN EP572088-A.  
PD 01-DEC-1993.  
XX  
PF 25-MAY-1993; 93EP-0201500.  
XX  
PR 25-MAY-1992; 92EP-0201492.  
XX  
PA (KONN ) GIST-BROCADES NV.  
PI Bonekamp AJ, Van Tilborg MWEM;  
XX  
DR WPT; 1993-379003/48.  
XX  
PT New mutant signal sequences with fewer potential cleavage sites -  
PT reduce processing ambiguity of signal peptidase(s)  
PS Claim 7; Page 15; 23pp; English.  
XX  
XX The two cleavage sites of B. licheniformis alpha-amylase signal  
CC peptide may give rise to two different mature proteins (AAR44429).  
CC To obtain a proper mature prod., the second recognition site  
CC has to be destroyed without interfering with the first recognition  
CC site. The combination of the B. licheniformis alpha-amylase  
CC signal sequence with human IL-3 shows a third possible cleavage  
CC site since mature IL-3 starts with Ala (AAR51699). In practice  
CC the third site is not used, as the signal peptidase is not  
CC active when the Ala-X-Ala is upstream of a pro.  
CC Mutant signal sequences with fewer potential cleavage sites  
CC reduce the processing ambiguity of signal peptidases.  
CC Mutated Bacillus alpha-amylase sequences in combination with  
CC IL-3 are illustrated in AAR44425-28.  
CC The oligonucleotide probes used to obtain the desired mutations  
CC are given in AAO52516-25.  
XX  
SQ Sequence 29 AA:

Query Match 0.8%; Score 7; DB 14; Length 29;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 KRUYARL 172  
|||||||  
DB 5 KRLUYARL 11

## RESULT 22

AAAR44428  
ID AAR44428 standard; Protein; 29 AA.

AC AAR44428;

DT 26-MAY-1994 (first entry)

DE Mutant alpha-amylase signal peptide.

KW Bacillus licheniformis; alpha-amylase; signal peptide;

KM peptidase; mature protein; recognition site; mutant;

XX cleavage site; interleukin; IL-3.

OS Bacillus licheniformis.

PN EP572088-A.

PD 01-DEC-1993.

PF 25-MAY-1993; 93EP-0201500.

PR 25-MAY-1992; 92EP-0201492.

PA (KONN) GIST-BROCADES NV.

PI Bonekamp AJ, Van Tilborg MWM;

DR WPI; 1993-379003/48.

PT New mutant signal sequences with fewer potential cleavage sites -

PS reduce processing ambiguity of signal peptidase(s)

XX Claim 7; Page 15; 23pp; English.

CC The two cleavage sites of B. licheniformis alpha-amylase signal

CC peptide may give rise to two different mature proteins (AAR44429).

CC To obtain a proper mature prod., the second recognition site

CC has to be destroyed without interfering with the first recognition

CC site. The combination of the B. licheniformis alpha-amylase

CC signal sequence with human IL-3 shows a third possible cleavage

CC site since mature IL-3 starts with Ala (AAR51699). In practice

CC the third site is not used, as the signal peptidase is not

CC active when the Ala-X-Ala is upstream of a Pro.

CC Mutant signal sequences with fewer potential cleavage sites

CC reduce the processing ambiguity of signal peptidases.

CC Mutated Bacillus alpha-amylase sequences in combination with

CC IL-3 are illustrated in AAR4425-28.

CC The oligonucleotide probes used to obtain the desired mutations

CC are given in AAR52516-25.

XX SQ Sequence 29 AA;

Query Match 0.8%; Score 7; DB 14; Length 29;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 KRLUYARL 172  
|||||||

DB 5 KRLUYARL 11

## RESULT 23

AAR74559  
ID AAR74559 standard; Peptide; 34 AA.

AC AAR74559;

DT 04-JAN-1996 (first entry)

DE Protease Nexin-1 N-terminal peptide.

KW Protease Nexin-1 N-terminal peptide; PN-1; variants;

KM protease inhibitors; urokinase inhibitor; inflammation;

KW tumour cell invasion; elastase inhibitor; emphysema;

XX septic shock; wound healing.

OS Homo sapiens.

PN W09511987-A1.

PD 04-MAY-1995.

PF 28-OCT-1994; 94WO-US11624.

PR 29-OCT-1993; 93US-0144758.

PA (INCY-) INCYTE PHARM INC.

PI Braxton SM, Scott RW;

DR WPI; 1995-17879/23.

PT New protease nexin-1 variant with altered active site - for treating

PT inflammation etc.; also proteins modified by polyethylene glycol

PT attached to Cys thiol residues, related DNA and nexin fusion

XX proteins.

XX Example A; Page 73; 122pp; English.

XX AAR74559 is the protease Nexin-1 (PN-1) N-terminal peptide. It was used

XX to isolate the complete PN-1 sequence, from which the active site

XX variants described in AAR74533-R74546 were derived. These variants have

XX different protease specificities and/or increased rate association

XX constants compared to the native PN-1. The variants can be used as

XX urokinase inhibitors (to treat inflammation and tumour cell invasion),

XX as elastase inhibitors (to treat emphysema and septic shock) and as

XX wound healing stimulators.

XX SQ Sequence 34 AA;

Query Match 0.8%; Score 7; DB 16; Length 34;

Best Local Similarity 100.0%; Pred. No. 61;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 PLSLLEEL 123  
|||||||

DB 5 PLSLLEEL 11

## RESULT 24

AAAW51936  
ID AAW51936 standard; Peptide; 34 AA.

AC AAW51936;

DT 26-OCT-1998 (first entry)

DE Human protease nexin I N-terminal peptide.

KW Nexin I type alpha; PN-1; serine protease inhibitor; serpin; human;

KM antiinflammatory; cysteine-PEGylated protein; polyethylene glycol;

XX PEG.

XX Homo sapiens.

PN US5766897-A.  
 XX  
 PD 16-JUN-1998.  
 XX  
 PF 21-JUN-1990; 90US-0542484.  
 XX  
 PR 21-APR-1995; 95US-0427100.  
 PR 21-JUN-1990; 90US-0542484.  
 PR 03-AUG-1992; 92US-0924294.  
 PR 29-OCT-1993; 93US-0144758.  
 PR 28-OCT-1994; 94WO-US11624.  
 XX  
 PA (INCY-) INCYTE PHARM INC.  
 XX  
 PI Braxton SM;  
 XX  
 DR WPI: 1998-361691/31.  
 XX  
 PT Chemically modified protease nexin-1 variants - comprise the amino  
 PT acid sequence of protease nexin-1 with certain residues replaced  
 PS with cysteine residues modified with polyethylene glycol  
 XX  
 PS Example A: Column 29; 34pp: English.  
 XX  
 CC This amino acid sequence comprises an N-terminal peptide of human  
 CC protease nexin-1 (PN-1, see also AA051933-34) isolated from human  
 CC forestin fibroblasts. Claimed chemically modified PN-1 variants  
 CC (see AA050329-56) comprise PN-1 polypeptides in which the amino acid  
 CC residues at positions adjacent to the reactive site centre are  
 CC replaced by corresponding amino acids of other serpins, and in  
 CC which at least cysteine residue of the polypeptide is coupled to  
 CC polyethylene glycol (PEG). Other claimed PEGylated PN-1 variants  
 CC have a substrate sequence of a different protease. PEG-modified  
 CC proteins have a prolonged half-life in vivo. They cause decreased  
 CC immunogenicity and antigenicity while retaining the same level of  
 CC biological activity as the naturally occurring protein. Specific  
 CC PN-1 variants are capable of inhibiting elastase in the presence of  
 CC heparin.  
 XX  
 SQ Sequence 34 AA;  
 XX  
 Query Match  
 Best Local Similarity 0.8%; Score 7; DB 19; Length 34;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 117 PLSTREL 123  
 DB 5 PLSTREL 11  
 XX  
 RESULT 25  
 AA088768  
 ID AA088768 standard; Peptide: 39 AA.  
 XX  
 AC AA088768;  
 XX  
 DT 18-JUN-2002 (first entry)  
 XX  
 DE Insulin/insulin-like growth factor receptor-binding peptide #724.  
 XX  
 KW Cytostatic; antidiabetic; neuroprotective; cerebroprotective;  
 KW ophthalmological; insulin; receptor; gene therapy; diabetes;  
 KW insulin-like growth factor-1; IGF-1; tumour; prostate; breast;  
 KW diabetic retinopathy; neurological diseases; stroke;  
 KW diabetic neuropathy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO20012771-A2.  
 XX  
 PD 04-OCT-2001.  
 XX  
 PF 29-MAR-2000; 2000WO-US08528.  
 XX

XX  
 PR 29-MAR-2000; 2000WO-US08528.  
 XX  
 PA (DGI B-) DGI BIOTECHNOLOGIES LLC.  
 PA (NOVO) NOVO NORDISK AS.  
 XX  
 PI Beasley J, Blume AJ, Schaeffer L, Pillutla R, Brandt J;  
 PI Brissette R, Spetzler J, Cheng W, Ostergaard S, Mandeckl WS;  
 PI Hansen PH, Ravera M, Hsiao K;  
 XX  
 DR WPI: 2002-025774/03.  
 XX  
 XX Modulating insulin activity in mammalian cells, for treating e.g.  
 PT diabetes and tumours, comprises using peptides that bind to insulin or  
 PT insulin-like growth factor receptors -  
 XX  
 PS Disclosure; Figure 1H; 390pp: English.  
 XX  
 CC The invention relates to a method of modulating insulin activity in  
 CC mammalian cells by administering a peptide that binds the insulin  
 CC receptor (IR). A composition containing a peptide, optionally expressed  
 CC from gene therapy vectors, that binds to site 1 of IR and an insulin  
 CC antagonist are useful for treating diabetes. Also, peptides that are  
 CC useful for treating insulin-like growth factor-1 (IGF-1) receptor are  
 CC (e.g. of prostate and breast) and diabetic retinopathy, while IGF-1  
 CC receptor agonists are useful for treating neurological diseases.  
 CC including stroke and diabetic neuropathy. The peptides are also useful in  
 CC screening for compounds that bind to IR or IGF-1 receptor, potential  
 CC and/or IGF-1 receptor-binding peptides and related amino acid sequences  
 CC of the invention.  
 XX  
 SQ Sequence 39 AA;  
 XX  
 Query Match  
 Best Local Similarity 0.8%; Score 7; DB 23; Length 39;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 523 ILGRISD 529  
 DB 6 ILGRISD 12  
 XX  
 RESULT 26  
 AAP60714  
 ID AAP60714 standard; Protein: 41 AA.  
 XX  
 AC AAP60714;  
 XX  
 DT 23-JUL-1991 (first entry)  
 XX  
 DE Sequence of the fusion of the alpha-amylase signal sequence to  
 DE alkaline phosphatase in the vector pNH218.  
 XX  
 KW Vector; Gram positive bacteria.  
 XX  
 OS Bacillus licheniformis.  
 XX  
 FH Key  
 FT Peptide 1..29 Location/Qualifiers  
 FT Region 30..36 /label= signal  
 FT /label= linker  
 FT Protein 37..41  
 FT /note= "starts at AA residue 5"  
 XX  
 PN WO8605812-A.  
 XX  
 PD 09-OCT-1986.  
 XX  
 PF 28-MAR-1986; 86WO-US00636.  
 XX



PR 28-MAR-1986; 86US-0845864.  
PR 29-MAR-1985; 85US-0717321.  
XX  
XX (BIOT-) BIOTECHNICA INT INC.  
PA  
PI Stephens MA, Rudolph CF, Hannett NM, Stassi DL, Pero JG;  
XX WPI; 1986-278825/42.  
DR N-PSDB; AAN60673.  
XX  
PT Vectors useful for transforming Gram positive bacteria - contg.  
PT secretory signal encoding sequence of Bacillus licheniformis  
PT alpha-amylase gene  
XX  
PS Example; Fig 8; 60pp; English.  
XX  
CC The vectors of the invention may be used to transform Gram positive  
CC bacteria for prodn. of polypeptides e.g. growth hormone. Using the  
CC vectors contg. a gene for alkaline phosphatase fused to a gene  
CC encoding growth hormone, expression can be monitored by observing  
CC enzymatic activity.  
XX  
SQ Sequence 41 AA;  
  
Query Match 0.8%; Score 7; DB 7; Length 41;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 166 KRLYARL 172  
|||||||  
Db 5 KRLYARL 11  
  
RESULT 27  
AAB65038  
ID AAB65038 standard; peptide: 41 AA.  
XX  
AC AAB65038;  
XX  
DT 23-MAR-2001 (first entry)  
XX  
DE Gene #5 associated peptide #6.  
XX  
KM Secreted protein; gene therapy; vaccine; cancer; leukemia;  
KM autoimmune disease; allergy; inflammation; graft rejection;  
KM hyperproliferation; cardiovascular; infection.  
XX  
OS Homo sapiens.  
XX  
XX W0200075375-A1.  
PN  
XX 14-DEC-2000.  
PD  
XX 02-JUN-2000; 2000WO-US15187.  
PF  
XX 07-JUN-1999; 99US-0137725.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX Ruben SM, Birse CE, Duan RD, Soppet DR, Rosen CA, Shi Y;  
PI Lafleur DW, Olsen HS, Edner R, Florence KA, Nl J, Young PE;  
XX WPI; 2001-061741/07.  
DR  
XX  
PT Nucleic acids encoding 26 human secreted polypeptides, useful e.g. for  
PT preventing, diagnosing and/or treating cancers and for promoting wound  
PT healing -  
XX  
PS Disclosure; Page 23; 530pp; English.  
XX  
CC The present invention relates to 26 secreted human proteins. The  
CC proteins may be used in the prevention, diagnosis and treatment of  
CC diseases associated with inappropriate polypeptide expression.

CC For example, they may be used in gene therapy or in vaccines.  
CC Typical of diseases which are potentially treatable are cancers  
CC (including leukemia), autoimmune diseases, allergies, inflammation,  
CC graft rejection, hyperproliferation, cardiovascular diseases  
CC (particularly critical limb ischemia and coronary disease) and any  
CC involving abnormal angiogenesis, neurodegeneration and/or  
CC infectious diseases.  
XX  
SQ Sequence 41 AA;  
  
Query Match 0.8%; Score 7; DB 22; Length 41;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 508 DRVLAIN 514  
|||||||  
Db 4 DRVLAIN 10  
  
RESULT 28  
AAP60715  
ID AAP60715 standard; Protein: 42 AA.  
XX  
AC AAP60715;  
XX  
DT 23-JUL-1991 (first entry)  
XX  
DE Sequence of the fusion of the alpha-amylase signal sequence to  
DE alkaline phosphatase in the vector p2/38.  
XX  
DE Vectro: Gram positive bacteria.  
XX  
XX Bacillus licheniformis.  
XX  
OS  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..29  
FT Region /label= signal  
FT /label= linker  
FT Protein 40..42  
FT /note= "starts at AA residue 12"  
XX  
XX W0605812-A.  
PN  
XX 09-OCT-1986.  
PD  
XX 28-MAR-1986; 86WO-US00636.  
PF  
XX 28-MAR-1986; 86US-0845864.  
PR 29-MAR-1985; 85US-0717321.  
XX  
PA (BIOT-) BIOTECHNICA INT INC.  
XX  
PI Stephens MA, Rudolph CF, Hannett NM, Stassi DL, Pero JG;  
XX WPI; 1986-278825/42.  
DR N-PSDB; AAN60674.  
XX  
PT Vectors useful for transforming Gram positive bacteria - contg.  
PT secretory signal encoding sequence of Bacillus licheniformis  
PT alpha-amylase gene  
XX  
PS Example; Fig 10; 60pp; English.  
XX  
CC The vectors of the invention may be used to transform Gram positive  
CC bacteria for prodn. of polypeptides e.g. growth hormone. Using the  
CC vectors contg. a gene for alkaline phosphatase fused to a gene  
CC encoding growth hormone, expression can be monitored by observing  
CC enzymatic activity.  
XX  
SQ Sequence 42 AA;  
  
Query Match 0.8%; Score 7; DB 7; Length 42;

Best Local Similarity 100.0%; Pred. No. 75;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 KRUYARL 172  
|||||  
DB 5 KRLYARL 11

## RESULT 29

AAV64841  
ID AAV64841 standard; Protein; 51 AA.

XX AAV64841;

DT 01-FEB-2000 (first entry)

DE Human 5' EST related polypeptide SEQ ID NO:1002.

XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;  
KW gene therapy; chromosome mapping; upstream regulatory sequence;  
KW forensic; location; development; protein synthesis; stability;  
KW regulation; identification.

XX Homo sapiens.

XX MO9953051-A2.

PD 21-OCT-1999.

PF 09-APR-1999; 99WO-1B00712.

XX 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

XX (GEST ) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-038446/03.

DR N-PSDB; AAZ42455.

XX Novel secreted protein 5' expressed sequence tag sequences used in  
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures  
XX  
PS Claim 3; Page 657-658; 837pp; English.

CC AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)  
CC sequences, corresponding to human secreted proteins. AAV64651 to  
CC AAV65438 represent the EST-related proteins corresponding to AAZ42265 to  
CC AAZ43052. The 5' ESTs can be used for producing secreted human gene  
CC products. They can be used to identify and isolate 5' untranslated  
CC regions (UTRs) and upstream regulatory regions which control the  
CC location, development stage, rate, and quantity of protein synthesis, as  
CC well as stability of mRNA. The ESTs are also useful as probes for  
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can  
CC also be used in forensic procedures to identify individuals, or in  
CC diagnostic procedures to identify individuals having genetic diseases  
CC resulting from abnormal gene expression. The products may also be used in  
CC gene therapy protocols. The nucleic acids encoding signal peptides can be  
CC used for directing extracellular secretion of a polypeptide or the  
CC insertion of a polypeptide into a membrane, or importing a polypeptide  
CC into a cell. The proteins encoded by the EST sequences may be useful in  
CC treating a variety of human conditions. Secreted proteins have  
CC therapeutic value, and the identification of new secreted proteins is  
CC valuable. AAZ42249 to AAZ42264 and AAV64644 to AAV64650 represent  
CC sequences used in the exemplification of the present invention.

XX Sequence 51 AA;

Query Match 0.8%; Score 7; DB 21; Length 51;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 LTDGILM 475  
|||||  
DB 2 LTDGILM 8

## RESULT 30

AAU22430  
ID AAU22430 standard; Protein; 51 AA.

XX AAU22430;

DT 18-DEC-2001 (first entry)

DE Human cardiovascular system antigen polypeptide SEQ ID No 1204.

XX Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;  
KW Chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;  
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;  
KW cerebroprotective; noctropic; antibacterial; virucide; fungicide; cancer;  
KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;  
KW hyperproliferative disorder; breast; liver; cardiovascular infection;  
KW cerebrovascular disorder; nervous system disorder; bacterial infection;  
KW fungal infection; viral infection; ocular disorder; endocrine disorder;  
KW gastrointestinal disorder; renal disorder; respiratory disorder;  
KW wound healing; skin aging; organ transplantation; tissue regeneration;  
KW anti-infertility.

XX Homo sapiens.

XX WO20015321-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01340.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218280.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 23-AUG-2000; 2000US-0227182.

PR 30-AUG-2000; 2000US-0227009.

PR 01-SEP-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000: 2000US-0229344.  
PR 01-SEP-2000: 2000US-0229345.  
PR 05-SEP-2000: 2000US-0229509.  
PR 05-SEP-2000: 2000US-0229513.  
PR 06-SEP-2000: 2000US-0230437.  
PR 06-SEP-2000: 2000US-0230438.  
PR 08-SEP-2000: 2000US-0231242.  
PR 08-SEP-2000: 2000US-0231243.  
PR 08-SEP-2000: 2000US-0231244.  
PR 08-SEP-2000: 2000US-0231413.  
PR 08-SEP-2000: 2000US-0231414.  
PR 08-SEP-2000: 2000US-0232080.  
PR 08-SEP-2000: 2000US-0232081.  
PR 12-SEP-2000: 2000US-0231968.  
PR 14-SEP-2000: 2000US-0232397.  
PR 14-SEP-2000: 2000US-0232398.  
PR 14-SEP-2000: 2000US-0232399.  
PR 14-SEP-2000: 2000US-0232400.  
PR 14-SEP-2000: 2000US-0232401.  
PR 14-SEP-2000: 2000US-0233063.  
PR 14-SEP-2000: 2000US-0233064.  
PR 14-SEP-2000: 2000US-0233065.  
PR 21-SEP-2000: 2000US-0234223.  
PR 21-SEP-2000: 2000US-0234274.  
PR 25-SEP-2000: 2000US-0234597.  
PR 25-SEP-2000: 2000US-0234598.  
PR 26-SEP-2000: 2000US-0235484.  
PR 27-SEP-2000: 2000US-0235834.  
PR 27-SEP-2000: 2000US-0235836.  
PR 29-SEP-2000: 2000US-0236327.  
PR 29-SEP-2000: 2000US-0236367.  
PR 29-SEP-2000: 2000US-0236368.  
PR 29-SEP-2000: 2000US-0236369.  
PR 29-SEP-2000: 2000US-0236370.  
PR 02-OCT-2000: 2000US-0236802.  
PR 02-OCT-2000: 2000US-0237037.  
PR 02-OCT-2000: 2000US-0237038.  
PR 02-OCT-2000: 2000US-0237039.  
PR 02-OCT-2000: 2000US-0237040.  
PR 13-OCT-2000: 2000US-0239935.  
PR 13-OCT-2000: 2000US-0239937.  
PR 20-OCT-2000: 2000US-0240960.  
PR 20-OCT-2000: 2000US-0241221.  
PR 20-OCT-2000: 2000US-024185.  
PR 20-OCT-2000: 2000US-0241786.  
PR 20-OCT-2000: 2000US-0241787.  
PR 20-OCT-2000: 2000US-0241808.  
PR 20-OCT-2000: 2000US-0241809.  
PR 20-OCT-2000: 2000US-0241826.  
PR 01-NOV-2000: 2000US-0244617.  
PR 08-NOV-2000: 2000US-0246474.  
PR 08-NOV-2000: 2000US-0246475.  
PR 08-NOV-2000: 2000US-0246476.  
PR 08-NOV-2000: 2000US-0246477.  
PR 08-NOV-2000: 2000US-0246478.  
PR 08-NOV-2000: 2000US-0246523.  
PR 08-NOV-2000: 2000US-0246524.  
PR 08-NOV-2000: 2000US-0246525.  
PR 08-NOV-2000: 2000US-0246526.  
PR 08-NOV-2000: 2000US-0246527.  
PR 08-NOV-2000: 2000US-0246528.  
PR 08-NOV-2000: 2000US-0246532.  
PR 08-NOV-2000: 2000US-0246609.  
PR 08-NOV-2000: 2000US-0246610.  
PR 08-NOV-2000: 2000US-0246611.  
PR 08-NOV-2000: 2000US-0246613.  
PR 17-NOV-2000: 2000US-0249207.  
PR 17-NOV-2000: 2000US-0249208.  
PR 17-NOV-2000: 2000US-0249209.  
PR 17-NOV-2000: 2000US-0249210.  
PR 17-NOV-2000: 2000US-0249211.  
PR 17-NOV-2000: 2000US-0249212.  
PR 17-NOV-2000: 2000US-0249213.

PR 17-NOV-2000: 2000US-0249214.  
PR 17-NOV-2000: 2000US-0249215.  
PR 17-NOV-2000: 2000US-0249216.  
PR 17-NOV-2000: 2000US-0249217.  
PR 17-NOV-2000: 2000US-0249218.  
PR 17-NOV-2000: 2000US-0249244.  
PR 17-NOV-2000: 2000US-0249245.  
PR 17-NOV-2000: 2000US-0249264.  
PR 17-NOV-2000: 2000US-0249267.  
PR 17-NOV-2000: 2000US-0249299.  
PR 17-NOV-2000: 2000US-0249300.  
PR 01-DEC-2000: 2000US-0250160.  
PR 01-DEC-2000: 2000US-0250391.  
PR 05-DEC-2000: 2000US-0251030.  
PR 05-DEC-2000: 2000US-0251988.  
PR 05-DEC-2000: 2000US-0256719.  
PR 06-DEC-2000: 2000US-0251479.  
PR 08-DEC-2000: 2000US-0251856.  
PR 08-DEC-2000: 2000US-0251858.  
PR 08-DEC-2000: 2000US-0251869.  
PR 08-DEC-2000: 2000US-0251989.  
PR 08-DEC-2000: 2000US-0251990.  
PR 11-DEC-2000: 2000US-0254097.  
PR 05-JAN-2001: 2001US-0259678.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-451930/48.  
DR N-PSDB: AAS35704.  
XX  
PT New cardiovascular system related polynucleotides and polypeptides,  
PT useful for diagnosing, treating and/or preventing disorders of the  
PT cardiovascular system -  
XX  
XX  
PS Claim 11: SEQ ID NO 1204: 674bp; English.  
XX  
CC Sequences AMU21852-AMU22466 represent the cardiovascular system antigen  
CC polypeptides of the invention. Cardiovascular system antigens and their  
CC associated polynucleotides are useful in the diagnosis, treatment and  
CC prevention of various types of disorders in e.g. humans, mice, rabbits,  
CC goats, horses, cats, dogs, chickens or sheep. A pathological condition  
CC can be determined by detecting the presence or absence of a mutation in a  
CC cardiovascular system antigen polynucleotide. The treatable disorders  
CC include autoimmune diseases such as rheumatoid arthritis,  
CC hyperproliferative disorders such as neoplasms of the breast or liver,  
CC cardiovascular disorders such as cardiac arrest, cerebrovascular  
CC disorders such as cerebral ischaemia, nervous system disorders such as  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,  
CC ocular disorders such as corneal infection, endocrine disorders such as  
CC premature labour and infertility, gastrointestinal disorders such as  
CC Crohn's disease, renal disorders such as glomerulonephritis and  
CC respiratory disorders such as asthma and pleurisy. The polypeptides can  
CC also be used to aid wound healing, to prevent skin aging due to sunburn,  
CC to maintain organs before transplantation, to regenerate tissues and in  
CC chemotaxis.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 0.8%; Score 7; DB 22; Length 51;

Best Local Similarity 100.0%; Pred. No. 89;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 LROGTALV 240

Db 25 LROGTALV 31

RESULT 31

AAM87002

ID AAM87002 standard; Protein: 62 AA.  
XX AAM87002;  
AC  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human Immune/Haematopoietic antigen SEQ ID NO:14595.  
XX  
KW Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;  
RV cytosolic; gene therapy; vaccine; metastasis.  
XX  
OS Homo sapiens.  
XX  
PN M0200157182-A2.  
PD  
XX 09-AUG-2001.  
PF  
XX 17-JAN-2001; 2001MO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217496.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225477.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226759.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 26-SEP-2000; 2000US-0234998.  
PR 27-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 29-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236369.  
PR 02-OCT-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 20-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.

```

PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM.
XX
XX WPI; 2001-483426/52.
XX
XX N-PSDB; AAK59783.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Claim 11; SEQ ID NO 14595; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK67694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 62 AA;
XX
XX Query Match 0.8%; Score 7; DB 22; Length 62;
XX Best Local Similarity 100.0%; Pred. No. 1.1e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 407 VRALSMD 413
XX |||||
DB 39 VRALSMD 45
XX
XX RESULT 32
XX AAY11419
XX ID AAY11419 standard; Protein: 64 AA.
XX
XX AAY11419;
XX
XX 21-JUN-1999 (first entry)
XX
XX Human 5' EST secreted protein SEQ ID NO 241.
XX
XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX forensic; gene therapy; chromosome mapping; signal peptide;
XX upstream regulatory sequence; cytokine activity; cell proliferation;
XX differentiation; haematopoiesis regulation; tissue growth regulation;
XX reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
XX thrombolytic; anti-inflammatory; tumour inhibition.
XX
XX Homo sapiens.
XX
XX OS
XX XX
XX PN WO9906551-A2.
XX
XX
XX

```

```

PD 11-FEB-1999.
XX
XX 31-JUL-1998; 98WO-1B01335.
XX
XX 01-AUG-1997; 97US-0905133.
XX
XX (GEST) GENSET.
XX
XX Ductert A, Dumas Milne Edwards J, Lacroix B;
XX
XX WPI; 1999-153781/13.
XX
XX N-PSDB; AAX39485.
XX
XX New nucleic acids encoding human secreted - proteins obtained from
XX cDNA libraries prepared from substantia nigra, cerebellum, surrenals
XX and fetal brain tissue
XX
XX Claim 34; Page 364; 434pp; English.
XX
XX AAX39440 to AAX39597 represent 5' expressed sequence tags (ESTs) for
XX human secreted proteins, and encode the proteins given in AAY11374 to
XX AAY11531, respectively. The proteins given represent the signal peptide
XX and an N-terminal fragment of a secreted protein. The nucleic acid
XX sequences can be used for producing secreted human gene products. They
XX can also be used to develop products for diagnosis and therapy. The
XX proteins obtained may have cytokine activity, cell
XX proliferation/differentiation activity, haematopoiesis regulating
XX activity, tissue growth regulating activity, reproductive hormone
XX regulating activity, chemotactic/chemokinetic activity, haemostatic and
XX thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, tumour inhibition activity or other activities. The products
XX can be used in forensic, gene therapy and chromosome mapping procedures.
XX The sequences can also be used for obtaining corresponding promoter
XX sequences. The nucleic acids encoding the signal peptide can be used for
XX directing extracellular secretion of a polypeptide or the insertion of a
XX polypeptide into a membrane, or importing a polypeptide into a cell.
XX
XX Sequence 64 AA;
XX
XX Query Match 0.8%; Score 7; DB 20; Length 64;
XX Best Local Similarity 100.0%; Pred. No. 1.1e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 573 LYVFVAS 579
XX |||||
DB 17 LYVFVAS 23
XX
XX RESULT 33
XX AAG77071
XX ID AAG77071 standard; Protein: 64 AA.
XX
XX AAG77071;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen protein SEQ ID NO:7835.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma.
XX
XX Homo sapiens.
XX
XX OS
XX XX
XX PN WO200122920-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
XX
XX 03-NOV-1999; 99US-0163280.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX

```



```

PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057669.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
PI Feng P, Ferlie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
PI Kyaw H, Lafleur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX
DR WPI: 1998-506364/43.
DR N-PSDB: AAV59785.
XX
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1; Page 694; 721pp; English.
XX
CC This sequence represents a secreted human protein encoded by the nucleic
CC acid molecule designated Gene 146 from the human cDNA clone HSNAK17
CC (deposited as clone ATCC 97903 and ATCC 209049).
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. AAV59502) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 186
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAV59511 for described uses).
XX
SQ Sequence 69 AA;
XX
Query Match 0.8%; Score 7; DB 19; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 388 ELLEQLL 394
DB 51 ELLEQLL 57
XX
RESULT 35
AAG01031
ID AAG01031 standard; Protein; 69 AA.
XX
AC AAG01031;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 5112.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX

```

```

XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PE 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GSEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI: 2000-500381/45.
DR N-PSDB: AAC01037.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13; SEQ ID 5112; 71pp + CD-ROM; English.
XX
CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
SQ Sequence 69 AA;
XX
Query Match 0.8%; Score 7; DB 21; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 482 FSASNLI 488
DB 59 FSASNLI 65
XX
RESULT 36
AAG01032
ID AAG01032 standard; Protein; 69 AA.
XX
AC AAG01032;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 5113.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PE 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GSEST ) GENSET.
XX

```

XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI: 2000-500381/45.  
DR N-PSDB: AAC01038.  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
PS Claim 13: SEQ ID 5113; 71bp + CD-ROM; English.  
XX  
CC The present sequence is a polypeptide encoded by one of a large number  
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
CC were prepared from total human RNAs or poly(A) RNAs derived from 30  
CC different tissues. EST sequences usually correspond mainly to the 3'  
CC untranslated region (UTR) of the mRNA because they are often obtained  
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
CC those cases where longer cDNA sequences have been obtained, the full 5'  
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
CC ends and can therefore be used to obtain full length cDNAs and genomic  
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
CC chromosome mapping procedures. They are used to obtain upstream  
CC regulatory sequences and to design expression and secretion vectors.  
XX  
SQ Sequence 69 AA:  
  
Query Match 0.8%; Score 7; DB 21; Length 69;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 482 FSASNLI 488  
Db 59 FSASNLI 65  
|||||||  
  
RESULT 37  
AAW74874  
ID AAW74874 standard; Protein: 70 AA.  
XX  
AC AAW74874;  
XX  
DT 25-JAN-1999 (first entry)  
XX  
DE Human secreted protein encoded by gene 146 clone HSNK17.  
XX  
DE Human; secreted protein; testis; tumour; foetal brain tissue;  
KW fusion protein; cancer; central nervous system; seizure;  
KW diagnosis; neurodegenerative disease.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 70 /label= unknown  
XX  
PN WC09839448-A2.  
XX  
PD 11-SEP-1998.  
XX  
PE 06-MAR-1998; 98MO-US04493.  
XX  
XX 02-OCT-1997; 97US-0061060.  
PR 07-MAR-1997; 97US-0038621.  
PR 07-MAR-1997; 97US-0040161.  
PR 07-MAR-1997; 97US-0040162.  
PR 07-MAR-1997; 97US-0040163.  
PR 07-MAR-1997; 97US-0040334.  
PR 07-MAR-1997; 97US-0040336.  
PR 07-MAR-1997; 97US-0040626.  
PR 11-APR-1997; 97US-0043311.

PR 11-APR-1997; 97US-0043312.  
PR 11-APR-1997; 97US-0043313.  
PR 11-APR-1997; 97US-0043314.  
PR 11-APR-1997; 97US-0043568.  
PR 11-APR-1997; 97US-0043569.  
PR 11-APR-1997; 97US-0043576.  
PR 11-APR-1997; 97US-0043578.  
PR 11-APR-1997; 97US-0043580.  
PR 11-APR-1997; 97US-0043669.  
PR 11-APR-1997; 97US-0043670.  
PR 11-APR-1997; 97US-0043671.  
PR 11-APR-1997; 97US-0043672.  
PR 11-APR-1997; 97US-0043674.  
PR 23-MAY-1997; 97US-0047492.  
PR 23-MAY-1997; 97US-0047500.  
PR 23-MAY-1997; 97US-0047501.  
PR 23-MAY-1997; 97US-0047502.  
PR 23-MAY-1997; 97US-0047503.  
PR 23-MAY-1997; 97US-0047504.  
PR 23-MAY-1997; 97US-0047581.  
PR 23-MAY-1997; 97US-0047582.  
PR 23-MAY-1997; 97US-0047583.  
PR 23-MAY-1997; 97US-0047584.  
PR 23-MAY-1997; 97US-0047585.  
PR 23-MAY-1997; 97US-0047586.  
PR 23-MAY-1997; 97US-0047587.  
PR 23-MAY-1997; 97US-0047588.  
PR 23-MAY-1997; 97US-0047589.  
PR 23-MAY-1997; 97US-0047590.  
PR 23-MAY-1997; 97US-0047592.  
PR 23-MAY-1997; 97US-0047593.  
PR 23-MAY-1997; 97US-0047594.  
PR 23-MAY-1997; 97US-0047595.  
PR 23-MAY-1997; 97US-0047596.  
PR 23-MAY-1997; 97US-0047597.  
PR 23-MAY-1997; 97US-0047598.  
PR 23-MAY-1997; 97US-0047599.  
PR 23-MAY-1997; 97US-0047600.  
PR 23-MAY-1997; 97US-0047601.  
PR 23-MAY-1997; 97US-0047612.  
PR 23-MAY-1997; 97US-0047613.  
PR 23-MAY-1997; 97US-0047614.  
PR 23-MAY-1997; 97US-0047615.  
PR 23-MAY-1997; 97US-0047616.  
PR 23-MAY-1997; 97US-0047617.  
PR 23-MAY-1997; 97US-0047618.  
PR 23-MAY-1997; 97US-0047632.  
PR 23-MAY-1997; 97US-0047633.  
PR 06-JUN-1997; 97US-0048964.  
PR 06-JUN-1997; 97US-0048965.  
PR 13-JUN-1997; 97US-0049610.  
PR 08-JUL-1997; 97US-0051926.  
PR 16-JUL-1997; 97US-0052874.  
PR 18-AUG-1997; 97US-0055274.  
PR 22-AUG-1997; 97US-0056630.  
PR 22-AUG-1997; 97US-0056631.  
PR 22-AUG-1997; 97US-0056632.  
PR 22-AUG-1997; 97US-0056636.  
PR 22-AUG-1997; 97US-0056637.  
PR 22-AUG-1997; 97US-0056662.  
PR 22-AUG-1997; 97US-0056664.  
PR 22-AUG-1997; 97US-0056665.  
PR 22-AUG-1997; 97US-0056845.  
PR 22-AUG-1997; 97US-0056862.  
PR 22-AUG-1997; 97US-0056864.  
PR 22-AUG-1997; 97US-0056872.  
PR 22-AUG-1997; 97US-0056874.  
PR 22-AUG-1997; 97US-0056875.  
PR 22-AUG-1997; 97US-0056876.  
PR 22-AUG-1997; 97US-0056877.  
PR 22-AUG-1997; 97US-0056878.  
PR 22-AUG-1997; 97US-0056879.  
PR 22-AUG-1997; 97US-0056880.  
PR 22-AUG-1997; 97US-0056881.  
PR 22-AUG-1997; 97US-0056882.  
PR 22-AUG-1997; 97US-0056884.



PR 22-AUG-1997; 97US-0056886.  
PR 22-AUG-1997; 97US-0056887.  
PR 22-AUG-1997; 97US-0056888.  
PR 22-AUG-1997; 97US-0056889.  
PR 22-AUG-1997; 97US-0056892.  
PR 22-AUG-1997; 97US-0056893.  
PR 22-AUG-1997; 97US-0056894.  
PR 22-AUG-1997; 97US-0056903.  
PR 22-AUG-1997; 97US-0056908.  
PR 22-AUG-1997; 97US-0056909.  
PR 22-AUG-1997; 97US-0056910.  
PR 22-AUG-1997; 97US-0056911.  
PR 05-SEP-1997; 97US-0057650.  
PR 05-SEP-1997; 97US-0057669.  
PR 05-SEP-1997; 97US-0057761.  
PR 12-SEP-1997; 97US-0058785.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;  
PI Feng P, Ferlie AM, Fischer CL, Florence KM, Greene JM, Hu JS;  
PI Kyaw H, Lafleur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;  
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
XX  
XX WPI: 1998-506364/43.  
DR N-PSDB: AAV59656.  
XX  
XX  
PT New isolated human genes and the secreted polypeptide(s) they encode  
PT - useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders  
XX  
XX  
PS Claim 1; Page 628; 721pp; English.  
XX  
XX This sequence represents a secreted human protein encoded by the nucleic  
CC acid molecule designated Gene 146 from the human cDNA clone HSNM17  
CC (deposited as clone ATCC 97903 and ATCC 209049).  
CC The gene can be used to generate fusion proteins by linking to the gene  
CC to a human immunoglobulin Fc portion (e.g. AAV59502) for increasing the  
CC stability of the fused protein as compared to the human protein only.  
CC The invention relates to 186 novel genes and their fragments (nucleic  
CC acid sequences: AAV59511-V59812; amino acid sequences AAW4731-W5026)  
CC which are useful for preventing, treating or ameliorating medical  
CC conditions e.g. by protein or gene therapy. Also, pathological  
CC conditions can be diagnosed by determining the amount of the new  
CC polypeptides in a sample or by determining the presence of mutations in  
CC the new polynucleotides. Specific uses are described for each of the 186  
CC polynucleotides, based on which tissues they are most highly expressed in  
CC (see AAV59511 for described uses).  
XX  
XX  
SQ Sequence 70 AA:  
OY  
Db 388 ELLEQL 394  
51 ELLEQL 57  
Query Match 0.8%; Score 7; DB 19; Length 70;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
RESULT 38  
ABP05664  
ID ABP05664 standard; Protein; 79 AA.  
XX  
XX ABP05664;  
XX  
XX 24-JUN-2002 (first entry)  
XX  
XX Human ORFX protein sequence SEQ ID NO:11310.  
XX  
XX Human: open reading frame; ORFX: gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW

KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.  
XX  
XX Homo sapiens.  
XX  
XX WO200192523-A2.  
XX  
XX 06-DEC-2001.  
XX  
XX  
XX 29-MAY-2001; 2001WO-US10836.  
XX  
XX 30-MAY-2000; 2000US-206132P.  
XX 29-AUG-2000; 2000US-228716P.  
XX  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Shinkets RA, Leach MD;  
XX  
XX WPI: 2002-106308/14.  
DR N-PSDB: ABN21416.  
XX  
XX  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders  
XX  
XX Disclosure; SEQ ID 11310; 1037pp; English.  
XX  
XX The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 79 AA:  
OY  
Db 387 RELLEQL 393  
56 RELLEQL 62  
Query Match 0.8%; Score 7; DB 23; Length 79;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
RESULT 39  
ABBI6701  
ID ABBI6701 standard; Protein; 82 AA.  
XX  
XX ABBI6701;  
XX

DT 23-JAN-2002 (first entry)  
XX  
DE Human nervous system related polypeptide SEQ ID NO 5358.  
XX  
XX Human: neurotropic; neuroprotective; cyostatic; dermatological; virucide;  
KW immunosuppressive; antineoplastic; anti-HIV; antibacterial; vulnery;  
KW antiparkinsonian; antischist; antianaemic; antiahrilic; cancer;  
KW antileptemic; hepatotropic; cerebroprotective; antileptemic;  
KW antileptemic; antidiabetic; antileptemic; antileptemic;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; nephrotropic; gene therapy; vaccine.  
OS Homo sapiens.  
PN WO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01334.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205467.  
PR 07-JUN-2000; 2000US-0209465.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225477.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 18-AUG-2000; 2000US-0225759.  
PR 22-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227709.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0231968.  
PR 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234597.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246539.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000: 2000US-0251988.  
 PR 05-DEC-2000: 2000US-0256719.  
 PR 06-DEC-2000: 2000US-0251479.  
 PR 08-DEC-2000: 2000US-0251856.  
 PR 08-DEC-2000: 2000US-0251868.  
 PR 08-DEC-2000: 2000US-0251869.  
 PR 08-DEC-2000: 2000US-0251989.  
 PR 08-DEC-2000: 2000US-0251990.  
 PR 11-DEC-2000: 2000US-0254097.  
 PR 05-JAN-2001: 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 DR WPI: 2001-541565/60.  
 DR N-PSDB: ABA13027.  
 XX  
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating nervous system  
 PT cancers and metastases -  
 XX  
 PS Claim 11: SEQ ID NO 5358; 1701bp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins  
 CC (AB14678-AB18001) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 SQ Sequence 82 AA:  
 QY 752 GSSGLVS 758  
 DB 69 GSSGLVS 75  
 Query Match 0.8%; Score 7; DB 22; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 752 GSSGLVS 758  
 DB 69 GSSGLVS 75  
 RESULT 40  
 AAG98900  
 ID AAG98900 standard; Protein; 82 AA.  
 AC AAG98900;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE E. coli growth and proliferation related protein sequence SEQ ID NO:370.  
 XX  
 KW Escherichia coli: growth; proliferation; microbial; antimicrobial;  
 KW bacterial infection; microorganism.  
 XX  
 OS Escherichia coli.  
 OS  
 PN WO200134810-A2.  
 XX  
 PD 17-MAY-2001.  
 XX

PF 09-NOV-2000: 2000WO-US30950.  
 XX  
 PR 09-NOV-1999: 99US-0164415.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Forsyth RA, Ohlsen K, Zyskind J;  
 DR WPI: 2001-335933/35.  
 DR N-PSDB: AAH84571.  
 XX  
 PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful  
 PT for screening for homologous genes and for designing expression vectors  
 PT -  
 XX  
 PS Claim 19: Page 442; 522pp; English.  
 XX  
 CC AAH84373 to AAH84499 represent Escherichia coli growth and proliferation  
 CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli  
 CC growth and proliferation related proteins given in AAG99078 and AAG98830  
 CC to AAG98999. (I) can be used as potential targets for the generation of  
 CC new antimicrobial agents, and for identification of compounds which  
 CC interact with the gene products of (I). In addition the expression of  
 CC (I) and the purification of the proteins, the purified proteins can be  
 CC used to generate reagents and screen small molecule libraries or other  
 CC candidate compound libraries for compounds that can be further developed  
 CC to yield novel antimicrobial compounds. In addition, nucleic acid probes  
 CC complementary to (I) that are specific for particular species of  
 CC microorganisms can be used to identify particular microorganism species  
 CC in clinical specimens, therefore, providing a rapid and dependable  
 CC method by which to identify the causative agents of a bacterial  
 CC infection. Also, antibodies generated against proteins translated from  
 CC mRNA transcribed from proliferation-required sequences can also be used  
 CC to screen for specific microorganisms that produce such proteins in a  
 CC species-specific manner. AAH84371 and AAH84670 represent sequencing  
 CC primers used in the isolation of E. coli growth and proliferation  
 CC related sequence, which are used in an example from the present  
 CC invention.  
 XX  
 SQ Sequence 82 AA:  
 QY 839 GQVLAAG 845  
 DB 7 GQVLAAG 13  
 Query Match 0.8%; Score 7; DB 22; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 839 GQVLAAG 845  
 DB 7 GQVLAAG 13

Search completed: April 28, 2003, 16:33:09  
 Job time : 97 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: April 28, 2003, 16:32:45 ; Search time 19 Seconds

(without alignments)  
1423.140 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 919  
Sequence: 1 MSKPYLFANRSEFPVALAA.....TGVKEGPNPIKHFITGTPF 919Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/prodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.0	851	1 US-08-369-796-2	Sequence 2, Appli
2	9	1.0	851	2 US-08-852-091-2	Sequence 2, Appli
3	9	1.0	851	2 US-08-820-754-2	Sequence 2, Appli
4	9	1.0	851	3 US-08-956-652-2	Sequence 2, Appli
5	9	1.0	851	3 US-08-956-869-2	Sequence 2, Appli
6	9	1.0	851	3 US-09-012-710-2	Sequence 2, Appli
7	9	1.0	851	3 US-08-948-547-2	Sequence 2, Appli
8	9	1.0	851	4 US-09-087-465-4	Sequence 2, Appli
9	9	1.0	851	4 US-09-364-970-2	Sequence 4, Appli
10	9	1.0	851	4 US-09-566-273-2	Sequence 2, Appli
11	9	1.0	851	4 US-08-956-653A-2	Sequence 2, Appli
12	9	1.0	851	5 PCT-US95-17025-2	Sequence 2, Appli
13	9	1.0	852	1 US-08-276-099A-13	Sequence 13, Appli
14	9	1.0	852	1 US-08-781-890-13	Sequence 13, Appli
15	8	0.9	582	4 US-09-242-632A-2	Sequence 2, Appli
16	7	0.8	29	1 US-07-960-510-4	Sequence 4, Appli
17	7	0.8	29	1 US-07-960-510-5	Sequence 5, Appli
18	7	0.8	29	1 US-07-960-510-6	Sequence 6, Appli
19	7	0.8	29	1 US-07-960-510-7	Sequence 7, Appli
20	7	0.8	34	6 5187089-8	Patent No. 5187089
21	7	0.8	34	6 5457090-5	Patent No. 5457090
22	7	0.8	34	6 5495001-16	Patent No. 5495001
23	7	0.8	35	6 5171673-7	Patent No. 5171673
24	7	0.8	69	4 US-09-149-476-601	Sequence 601, App
25	7	0.8	70	4 US-09-149-476-465	Sequence 465, App
26	7	0.8	95	1 US-08-438-753B-24	Sequence 24, Appli
27	7	0.8	95	1 US-08-443-883A-24	Sequence 24, Appli

28	7	0.8	95	2 US-08-631-328-24	Sequence 24, Appli
29	7	0.8	95	2 US-08-455-524B-24	Sequence 24, Appli
30	7	0.8	95	2 US-08-455-021B-24	Sequence 24, Appli
31	7	0.8	159	3 US-09-045-467-24	Sequence 2, Appli
32	7	0.8	175	2 US-08-796-792-2	Sequence 26, Appli
33	7	0.8	175	2 US-08-245-511-26	Sequence 12, Appli
34	7	0.8	175	2 US-08-600-993A-26	Sequence 12, Appli
35	7	0.8	264	4 US-09-323-872A-12	Sequence 64, Appli
36	7	0.8	349	2 US-08-724-864-64	Sequence 2, Appli
37	7	0.8	350	4 US-09-134-001C-5150	Sequence 5150, Ap
38	7	0.8	355	4 US-09-325-932A-191	Sequence 191, App
39	7	0.8	397	6 5187089-2	Patent No. 5187089
40	7	0.8	397	6 5187089-9	Patent No. 5187089
41	7	0.8	397	6 5187089-10	Patent No. 5187089
42	7	0.8	397	6 5187089-11	Patent No. 5187089
43	7	0.8	397	6 5187089-12	Patent No. 5187089
44	7	0.8	397	6 5457090-2	Patent No. 5457090
45	7	0.8	397	6 5457090-4	Patent No. 5457090
46	7	0.8	397	6 5495001-7	Patent No. 5495001
47	7	0.8	398	6 5187089-4	Patent No. 5187089
48	7	0.8	398	6 5187089-5	Patent No. 5187089
49	7	0.8	398	6 5187089-6	Patent No. 5187089
50	7	0.8	398	6 5187089-7	Patent No. 5187089
51	7	0.8	398	6 5495001-9	Patent No. 5495001
52	7	0.8	434	1 US-08-337-602-2	Sequence 2, Appli
53	7	0.8	434	3 US-08-558-135-2	Sequence 1, Appli
54	7	0.8	488	4 US-08-985-343-1	Sequence 35, Appli
55	7	0.8	511	1 US-08-468-700-35	Sequence 3, Appli
56	7	0.8	511	1 US-08-645-971-3	Sequence 33, Appli
57	7	0.8	511	2 US-08-468-220-33	Sequence 33, Appli
58	7	0.8	511	2 US-08-468-698-33	Sequence 35, Appli
59	7	0.8	511	2 US-08-704-706A-35	Sequence 2, Appli
60	7	0.8	511	3 US-08-890-383-4	Sequence 4, Appli
61	7	0.8	511	3 US-08-890-383-4	Sequence 2, Appli
62	7	0.8	511	3 US-08-914-679A-2	Sequence 4, Appli
63	7	0.8	511	3 US-08-914-679A-4	Sequence 4, Appli
64	7	0.8	511	4 US-08-985-659-36	Sequence 36, Appli
65	7	0.8	511	4 US-08-194-664A-33	Sequence 33, Appli
66	7	0.8	511	5 PCT-US94-01553A-33	Sequence 33, Appli
67	7	0.8	511	5 PCT-US95-10426-33	Sequence 33, Appli
68	7	0.8	512	1 US-07-623-953-5	Sequence 5, Appli
69	7	0.8	512	1 US-08-720-899-2	Sequence 2, Appli
70	7	0.8	512	1 US-08-459-610-2	Sequence 2, Appli
71	7	0.8	512	2 US-08-343-804-2	Sequence 2, Appli
72	7	0.8	512	2 US-08-687-399-2	Sequence 2, Appli
73	7	0.8	512	2 US-08-600-908A-2	Sequence 2, Appli
74	7	0.8	512	3 US-08-683-838A-2	Sequence 2, Appli
75	7	0.8	512	4 US-08-985-659-34	Sequence 34, Appli
76	7	0.8	516	4 US-09-636-252A-2	Sequence 2, Appli
77	7	0.8	521	1 US-08-868-373-14	Sequence 14, Appli
78	7	0.8	521	2 US-08-915-003-7	Sequence 7, Appli
79	7	0.8	534	4 US-08-642-247-7	Sequence 7, Appli
80	7	0.8	534	4 US-09-199-637A-67	Sequence 67, Appli
81	7	0.8	551	4 US-09-615-192A-348	Sequence 348, App
82	7	0.8	560	3 US-08-814-052-6	Sequence 6, Appli
83	7	0.8	612	2 US-08-812-829-6	Sequence 11, Appli
84	7	0.8	612	4 US-08-752-307B-11	Sequence 11, Appli
85	7	0.8	612	4 US-09-991-326-11	Sequence 11, Appli
86	7	0.8	631	3 US-08-814-052-8	Sequence 8, Appli
87	7	0.8	631	3 US-08-812-829-8	Sequence 8, Appli
88	7	0.8	655	4 US-09-228-986-70	Sequence 70, Appli
89	7	0.8	719	4 US-09-708-426-4	Sequence 4, Appli
90	7	0.8	750	3 US-08-814-052-2	Sequence 2, Appli
91	7	0.8	750	3 US-08-812-829-2	Sequence 2, Appli
92	7	0.8	1050	3 US-09-045-632-19	Sequence 49, Appli
93	7	0.8	1050	3 US-09-045-632-50	Sequence 50, Appli
94	7	0.8	1066	4 US-08-390-874C-11	Sequence 11, Appli
95	7	0.8	1066	4 US-09-265-772-11	Sequence 11, Appli
96	7	0.8	1069	2 US-08-162-081B-37	Sequence 37, Appli
97	7	0.8	1069	2 US-08-780-872-37	Sequence 37, Appli
98	7	0.8	1069	4 US-09-085-957-37	Sequence 36, Appli
99	7	0.8	1080	2 US-08-162-081B-36	Sequence 36, Appli
100	7	0.8			



247	6	0.7	205	4	US-09-724-864-37	Sequence 37, Appl	320	6	0.7	309	4	US-09-205-426-118	Sequence 118, App
248	6	0.7	207	3	US-08-957-302A-10	Sequence 10, Appl	321	6	0.7	309	4	US-09-651-200-22	Sequence 22, Appl
249	6	0.7	207	4	US-09-542-403-10	Sequence 10, Appl	322	6	0.7	309	5	PCT-US95-02576-21	Sequence 21, Appl
250	6	0.7	210	3	US-08-985-526-21	Sequence 21, Appl	323	6	0.7	311	2	US-08-318-837-9	Sequence 9, Appl
251	6	0.7	214	4	US-09-587-066-6	Sequence 6, Appl	324	6	0.7	312	4	US-08-961-083-12	Sequence 12, Appl
252	6	0.7	214	4	US-08-999-774A-4	Sequence 4, Appl	325	6	0.7	313	3	US-09-362-506-2	Sequence 2, Appl
253	6	0.7	221	4	US-09-382-155-17	Sequence 17, Appl	326	6	0.7	314	4	US-08-205-697A-13	Sequence 13, Appl
254	6	0.7	221	4	US-09-074-044A-17	Sequence 17, Appl	327	6	0.7	314	4	US-08-702-525-13	Sequence 13, Appl
255	6	0.7	222	2	US-08-090-048-1	Sequence 1, Appl	328	6	0.7	314	5	PCT-US95-02576-13	Sequence 13, Appl
256	6	0.7	222	2	US-08-292-550-1	Sequence 1, Appl	329	6	0.7	315	1	US-07-729-099-1	Sequence 1, Appl
257	6	0.7	222	2	US-07-927-661A-1	Sequence 1, Appl	330	6	0.7	315	1	US-08-257-392-1	Sequence 1, Appl
258	6	0.7	223	2	US-08-121-436A-2	Sequence 2, Appl	331	6	0.7	315	3	US-08-770-035-1	Sequence 1, Appl
259	6	0.7	223	4	US-09-254-733-7	Sequence 7, Appl	332	6	0.7	317	1	US-07-688-299-1	Sequence 1, Appl
260	6	0.7	225	2	US-08-886-765-2	Sequence 2, Appl	333	6	0.7	317	1	US-07-980-517A-1	Sequence 1, Appl
261	6	0.7	225	4	US-09-115-660-2	Sequence 2, Appl	334	6	0.7	318	1	US-07-688-299-3	Sequence 3, Appl
262	6	0.7	228	4	US-09-724-864-46	Sequence 46, Appl	335	6	0.7	318	1	US-07-688-299-13	Sequence 13, Appl
263	6	0.7	229	4	US-09-604-978-9	Sequence 9, Appl	336	6	0.7	321	1	US-08-362-670B-26	Sequence 26, Appl
264	6	0.7	242	2	US-08-622-352A-3	Sequence 3, Appl	337	6	0.7	321	1	US-08-333-576C-26	Sequence 26, Appl
265	6	0.7	242	2	US-08-826-390-3	Sequence 3, Appl	338	6	0.7	321	5	US-08-808-324-26	Sequence 26, Appl
266	6	0.7	250	2	US-08-612-788-29	Sequence 29, Appl	339	6	0.7	321	5	PCT-US94-14030A-26	Sequence 26, Appl
267	6	0.7	250	2	US-08-612-788-33	Sequence 33, Appl	340	6	0.7	323	4	US-09-462-844-4	Sequence 4, Appl
268	6	0.7	250	3	US-09-066-028-29	Sequence 29, Appl	341	6	0.7	325	1	US-08-276-919-4	Sequence 4, Appl
269	6	0.7	250	3	US-09-066-028-33	Sequence 33, Appl	342	6	0.7	325	1	US-08-276-919-13	Sequence 13, Appl
270	6	0.7	259	4	US-08-961-083-174	Sequence 174, App	343	6	0.7	325	1	US-08-776-088-4	Sequence 4, Appl
271	6	0.7	259	4	US-09-261-599B-3	Sequence 3, Appl	344	6	0.7	325	1	US-08-776-088-18	Sequence 18, Appl
272	6	0.7	259	4	US-09-456-455A-3	Sequence 3, Appl	345	6	0.7	325	3	US-08-581-148C-4	Sequence 4, Appl
273	6	0.7	266	4	US-09-199-637A-157	Sequence 157, App	346	6	0.7	325	5	PCT-US95-09145A-4	Sequence 4, Appl
274	6	0.7	266	4	US-09-134-001C-5453	Sequence 5453, Ap	347	6	0.7	325	5	PCT-US95-09145A-18	Sequence 18, Appl
275	6	0.7	271	1	US-08-276-919-10	Sequence 10, Appl	348	6	0.7	327	2	US-08-997-080-162	Sequence 162, App
276	6	0.7	271	1	US-08-776-088-13	Sequence 13, Appl	349	6	0.7	327	2	US-08-997-362-162	Sequence 162, App
277	6	0.7	271	5	PCT-US95-09145A-13	Sequence 15, Appl	350	6	0.7	327	4	US-09-095-855-162	Sequence 162, App
278	6	0.7	274	1	US-08-776-088-15	Sequence 15, Appl	351	6	0.7	327	4	US-09-324-542-162	Sequence 162, App
279	6	0.7	274	2	US-08-978-404B-5	Sequence 5, Appl	352	6	0.7	327	4	US-09-134-001C-3511	Sequence 3511, Ap
280	6	0.7	274	5	PCT-US95-09145A-15	Sequence 15, Appl	353	6	0.7	327	4	US-09-205-426-162	Sequence 162, App
281	6	0.7	278	4	US-09-060-410-17	Sequence 17, Appl	354	6	0.7	328	3	US-08-486-099-92	Sequence 92, Appl
282	6	0.7	281	4	US-09-134-001C-3491	Sequence 3491, Ap	355	6	0.7	328	3	US-08-360-107A-102	Sequence 102, App
283	6	0.7	282	1	US-08-324-301-15	Sequence 15, Appl	356	6	0.7	328	3	US-08-484-223B-92	Sequence 92, Appl
284	6	0.7	282	5	PCT-US94-09752-3	Sequence 3, Appl	357	6	0.7	328	3	US-08-919-597-92	Sequence 92, Appl
285	6	0.7	291	1	US-08-358-171-2	Sequence 2, Appl	358	6	0.7	328	3	US-08-475-668A-92	Sequence 92, Appl
286	6	0.7	291	3	US-09-090-947-2	Sequence 2, Appl	359	6	0.7	328	3	US-08-485-551A-92	Sequence 92, Appl
287	6	0.7	291	4	US-09-275-742-2	Sequence 2, Appl	360	6	0.7	328	3	US-08-471-913A-92	Sequence 92, Appl
288	6	0.7	293	4	US-09-134-001C-5333	Sequence 5333, Ap	361	6	0.7	328	3	US-08-485-264A-92	Sequence 92, Appl
289	6	0.7	294	4	US-08-973-334-2	Sequence 2, Appl	362	6	0.7	328	4	US-08-474-349A-92	Sequence 92, Appl
290	6	0.7	294	4	US-09-563-869A-2	Sequence 2, Appl	363	6	0.7	328	4	US-08-255-208A-28	Sequence 28, Appl
291	6	0.7	294	4	US-08-549-489-2	Sequence 2, Appl	364	6	0.7	330	4	US-09-134-001C-3582	Sequence 3582, Ap
292	6	0.7	297	2	US-09-027-013-3	Sequence 3, Appl	365	6	0.7	330	4	US-09-134-001C-3582	Sequence 3582, Ap
293	6	0.7	297	3	US-09-173-581-1	Sequence 1, Appl	366	6	0.7	331	3	US-08-356-419-25	Sequence 25, Appl
294	6	0.7	297	3	US-09-244-233-3	Sequence 3, Appl	367	6	0.7	333	4	US-09-097-319A-2	Sequence 2, Appl
295	6	0.7	297	4	US-08-845-258-26	Sequence 26, Appl	368	6	0.7	334	2	US-08-933-750C-16	Sequence 16, Appl
296	6	0.7	297	4	US-08-990-571-26	Sequence 26, Appl	369	6	0.7	334	4	US-09-234-613-16	Sequence 16, Appl
297	6	0.7	297	4	US-09-420-915-1	Sequence 1, Appl	370	6	0.7	336	4	US-09-457-040B-21	Sequence 21, Appl
298	6	0.7	297	4	US-08-723-142A-26	Sequence 26, Appl	371	6	0.7	337	4	US-09-134-001C-3799	Sequence 3799, Ap
299	6	0.7	297	4	US-09-528-784A-26	Sequence 26, Appl	372	6	0.7	337	1	US-08-248-629A-2	Sequence 2, Appl
300	6	0.7	299	4	US-09-134-001C-5227	Sequence 5227, Ap	373	6	0.7	339	1	US-08-248-629A-6	Sequence 6, Appl
301	6	0.7	300	4	US-09-284-665A-10	Sequence 10, Appl	374	6	0.7	339	1	US-08-451-932-2	Sequence 2, Appl
302	6	0.7	301	4	US-09-288-143-98	Sequence 98, Appl	375	6	0.7	339	1	US-08-451-932-6	Sequence 6, Appl
303	6	0.7	303	4	US-08-158-735A-14	Sequence 14, Appl	376	6	0.7	339	1	US-08-452-260-2	Sequence 2, Appl
304	6	0.7	303	4	US-08-818-112-92	Sequence 92, Appl	377	6	0.7	339	1	US-08-452-260-6	Sequence 6, Appl
305	6	0.7	303	4	US-08-818-111-93	Sequence 93, Appl	378	6	0.7	339	1	US-08-326-785-2	Sequence 2, Appl
306	6	0.7	303	4	US-09-056-556-92	Sequence 92, Appl	379	6	0.7	339	1	US-08-326-785-6	Sequence 6, Appl
307	6	0.7	303	4	US-09-651-200-23	Sequence 23, Appl	380	6	0.7	339	2	US-08-612-788-2	Sequence 2, Appl
308	6	0.7	303	4	US-09-072-596-93	Sequence 93, Appl	381	6	0.7	339	2	US-08-612-788-6	Sequence 6, Appl
309	6	0.7	304	4	US-09-134-001C-3492	Sequence 3492, Ap	382	6	0.7	339	2	US-08-605-598B-2	Sequence 2, Appl
310	6	0.7	306	4	US-09-342-647-26	Sequence 26, Appl	383	6	0.7	339	2	US-08-605-598B-6	Sequence 6, Appl
311	6	0.7	309	2	US-08-456-104-4	Sequence 4, Appl	384	6	0.7	339	2	US-08-429-743-2	Sequence 2, Appl
312	6	0.7	309	2	US-08-997-080-118	Sequence 118, App	385	6	0.7	339	2	US-08-429-743-6	Sequence 6, Appl
313	6	0.7	309	3	US-08-997-362-118	Sequence 118, App	386	6	0.7	339	2	US-08-866-735-2	Sequence 2, Appl
314	6	0.7	309	3	US-08-479-744A-23	Sequence 23, Appl	387	6	0.7	339	2	US-08-866-735-6	Sequence 6, Appl
315	6	0.7	309	4	US-08-280-757B-23	Sequence 23, Appl	388	6	0.7	339	3	US-09-066-028-2	Sequence 2, Appl
316	6	0.7	309	4	US-09-095-855-118	Sequence 118, App	389	6	0.7	339	3	US-09-066-028-6	Sequence 6, Appl
317	6	0.7	309	4	US-08-305-697A-21	Sequence 21, Appl	390	6	0.7	339	5	PCT-US95-05107-2	Sequence 2, Appl
318	6	0.7	309	4	US-08-702-525-21	Sequence 21, Appl	391	6	0.7	339	5	PCT-US95-05107-6	Sequence 6, Appl
319	6	0.7	309	4	US-09-324-542-118	Sequence 118, App	392	6	0.7	341	1	US-08-396-957A-5	Sequence 5, Appl

393	6	0.7	341	3	US-08-389-812-4	Sequence 4, Appl	466	6	0.7	383	2	US-08-760-936-4	Sequence 4, Appl
394	6	0.7	341	3	US-09-270-813-4	Sequence 4, Appl	467	6	0.7	384	2	US-08-935-887-1	Sequence 1, Appl
395	6	0.7	341	3	US-08-591-605-6	Sequence 6, Appl	468	6	0.7	386	1	US-08-758-213-1	Sequence 1, Appl
396	6	0.7	343	4	US-08-279-058B-9	Sequence 9, Appl	469	6	0.7	386	2	US-08-692-787-48	Sequence 11, Appl
397	6	0.7	343	4	US-08-828-323-9	Sequence 9, Appl	470	6	0.7	386	4	US-09-176-666-11	Sequence 48, Appl
398	6	0.7	344	4	US-09-134-001C-5004	Sequence 5004, Ap	471	6	0.7	386	4	US-09-176-666-11	Sequence 10, Appl
399	6	0.7	345	4	US-09-364-230-8	Sequence 8, Appl	472	6	0.7	387	4	US-09-097-199-48	Sequence 48, Appl
400	6	0.7	346	2	US-08-702-153-2	Sequence 2, Appl	473	6	0.7	388	4	US-09-176-666-9	Sequence 9, Appl
401	6	0.7	346	2	US-08-802-153-4	Sequence 4, Appl	474	6	0.7	389	4	US-09-176-666-8	Sequence 8, Appl
402	6	0.7	346	2	US-08-702-153-2	Sequence 4, Appl	475	6	0.7	389	4	US-09-134-001C-3161	Sequence 3161, Ap
403	6	0.7	351	1	US-08-468-847B-16	Sequence 16, Appl	476	6	0.7	390	4	US-09-176-666-7	Sequence 7, Appl
404	6	0.7	352	1	US-08-196-989B-2	Sequence 39, Appl	477	6	0.7	391	4	US-09-123-912-110	Sequence 110, App
405	6	0.7	352	2	US-08-612-788-39	Sequence 2, Appl	478	6	0.7	391	4	US-09-643-597-110	Sequence 110, App
406	6	0.7	352	2	US-08-760-936-2	Sequence 39, Appl	479	6	0.7	391	4	US-09-643-597-110	Sequence 6, Appl
407	6	0.7	352	3	US-09-066-028-39	Sequence 2, Appl	480	6	0.7	392	2	US-08-387-942C-23	Sequence 23, Appl
408	6	0.7	354	1	US-08-276-919-2	Sequence 2, Appl	481	6	0.7	392	4	US-08-176-666-5	Sequence 5, Appl
409	6	0.7	354	1	US-08-776-088-2	Sequence 6, Appl	482	6	0.7	393	4	US-09-176-666-4	Sequence 4, Appl
410	6	0.7	354	1	US-08-776-088-2	Sequence 2, Appl	483	6	0.7	393	6	5512669-6	Patent No. 5512669
411	6	0.7	354	3	US-09-325-320-2	Sequence 2, Appl	484	6	0.7	393	6	5512669-6	Patent No. 5512669
412	6	0.7	354	3	US-09-585-109-2	Sequence 2, Appl	485	6	0.7	394	4	US-09-176-666-3	Sequence 3, Appl
413	6	0.7	354	5	PCT-US95-09145A-2	Sequence 2, Appl	486	6	0.7	395	1	US-07-931-943-2	Sequence 2, Appl
414	6	0.7	354	4	PCT-US95-09145A-6	Sequence 6, Appl	487	6	0.7	395	2	US-08-624-601-15	Sequence 15, Appl
415	6	0.7	355	1	US-08-153-848-28	Sequence 28, Appl	488	6	0.7	396	4	US-09-176-666-2	Sequence 2, Appl
416	6	0.7	355	1	US-08-153-848-32	Sequence 32, Appl	489	6	0.7	396	4	US-09-948-997-5	Sequence 5, Appl
417	6	0.7	355	3	US-09-299-843A-28	Sequence 28, Appl	490	6	0.7	397	3	US-08-348-817A-5	Sequence 5, Appl
418	6	0.7	355	4	US-09-299-843A-32	Sequence 32, Appl	491	6	0.7	397	4	US-09-463-238-15	Sequence 15, Appl
419	6	0.7	355	4	US-09-088-337B-28	Sequence 28, Appl	492	6	0.7	398	3	US-09-189-035-6	Sequence 6, Appl
420	6	0.7	355	4	US-09-088-337B-28	Sequence 32, Appl	493	6	0.7	398	3	US-08-928-881-26	Sequence 26, Appl
421	6	0.7	355	4	US-09-088-337B-32	Sequence 32, Appl	494	6	0.7	398	4	US-09-382-086-6	Sequence 6, Appl
422	6	0.7	355	5	PCT-US93-11153-28	Sequence 28, Appl	495	6	0.7	398	4	US-09-176-666-1	Sequence 1, Appl
423	6	0.7	355	5	PCT-US93-11153-32	Sequence 32, Appl	496	6	0.7	398	4	US-09-543-921-26	Sequence 26, Appl
424	6	0.7	357	1	US-08-468-847B-17	Sequence 17, Appl	497	6	0.7	398	4	US-09-266-014-4	Sequence 4, Appl
425	6	0.7	357	1	US-09-253-316-25	Sequence 25, Appl	498	6	0.7	399	4	US-08-765-907A-10	Sequence 10, Appl
426	6	0.7	358	1	US-08-034-650-10	Sequence 10, Appl	499	6	0.7	399	4	US-09-134-001C-4571	Sequence 4571, Ap
427	6	0.7	358	1	US-08-449-015-10	Sequence 10, Appl	500	6	0.7	400	1	US-07-730-953-2	Sequence 1, Appl
428	6	0.7	358	2	US-08-558-823-19	Sequence 19, Appl	501	6	0.7	400	1	US-08-961-083-190	Sequence 19, Appl
429	6	0.7	358	2	US-08-558-823-19	Sequence 2, Appl	502	6	0.7	400	4	US-09-123-912-112	Sequence 112, App
430	6	0.7	358	3	US-09-116-115-2	Sequence 2, Appl	503	6	0.7	400	4	US-09-643-597-112	Sequence 112, App
431	6	0.7	358	4	US-09-541-762-2	Sequence 2, Appl	504	6	0.7	402	4	US-09-643-597-112	Sequence 5, Appl
432	6	0.7	359	4	US-08-809-103B-2	Sequence 4, Appl	505	6	0.7	402	4	US-09-029-755C-5	Sequence 5, Appl
433	6	0.7	359	4	US-08-809-103B-4	Sequence 4, Appl	506	6	0.7	403	2	US-08-533-669A-10	Sequence 2, Appl
434	6	0.7	359	4	US-08-809-103B-6	Sequence 6, Appl	507	6	0.7	403	2	US-08-607-509-2	Sequence 2, Appl
435	6	0.7	359	4	US-08-809-103B-8	Sequence 8, Appl	508	6	0.7	403	2	US-08-607-509-4	Sequence 4, Appl
436	6	0.7	359	4	US-09-625-188-8	Sequence 8, Appl	509	6	0.7	403	2	US-08-454-036-2	Sequence 2, Appl
437	6	0.7	362	4	US-09-134-001C-5403	Sequence 5403, Ap	510	6	0.7	403	2	US-08-634-642-4	Sequence 4, Appl
438	6	0.7	363	1	US-08-022-411-2	Sequence 13, Appl	511	6	0.7	403	2	US-08-634-642-4	Sequence 4, Appl
439	6	0.7	363	1	US-08-984-618-13	Sequence 4, Appl	512	6	0.7	403	3	US-08-989-370-2	Sequence 2, Appl
440	6	0.7	368	1	US-08-093-372-4	Sequence 2, Appl	513	6	0.7	403	3	US-08-989-370-4	Sequence 4, Appl
441	6	0.7	368	1	US-09-000-092-2	Sequence 4, Appl	514	6	0.7	403	4	US-09-183-861-10	Sequence 10, Appl
442	6	0.7	368	4	US-09-000-092-4	Sequence 6, Appl	515	6	0.7	403	5	PCT-US95-05064-2	Patent No. 5212296
443	6	0.7	368	4	US-09-000-092-6	Sequence 8, Appl	516	6	0.7	403	5	PCT-US95-05064-2	Patent No. 5212296
444	6	0.7	368	4	US-09-000-092-8	Sequence 10, Appl	517	6	0.7	405	2	US-08-519-081-2	Sequence 2, Appl
445	6	0.7	368	4	US-09-000-092-10	Sequence 41, Appl	518	6	0.7	405	2	US-08-519-081-2	Sequence 2, Appl
446	6	0.7	368	4	US-09-413-574-14	Sequence 14, Appl	519	6	0.7	405	2	US-08-754-356C-2	Sequence 2, Appl
447	6	0.7	374	1	US-08-095-726-14	Sequence 14, Appl	520	6	0.7	409	4	US-09-029-755C-2	Sequence 2, Appl
448	6	0.7	374	1	US-08-096-623A-14	Sequence 14, Appl	521	6	0.7	411	1	US-08-399-561-2	Sequence 2, Appl
449	6	0.7	375	1	US-08-121-714-2	Sequence 5, Appl	522	6	0.7	412	4	US-09-445-472-1	Sequence 4, Appl
450	6	0.7	375	1	US-08-736-361A-5	Sequence 2, Appl	523	6	0.7	414	2	US-08-305-764C-58	Sequence 58, Appl
451	6	0.7	375	1	US-08-477-108A-2	Sequence 2, Appl	524	6	0.7	418	2	US-08-305-764C-58	Sequence 58, Appl
452	6	0.7	375	2	US-08-477-112-2	Sequence 2, Appl	525	6	0.7	422	4	US-09-427-501-2	Sequence 2, Appl
453	6	0.7	375	5	PCT-US93-08322-2	Sequence 41, Appl	526	6	0.7	423	3	US-08-985-526-23	Sequence 23, Appl
454	6	0.7	378	2	US-08-612-788-41	Sequence 41, Appl	527	6	0.7	428	6	5179023-4	Patent No. 5179023
455	6	0.7	381	2	US-09-066-028-41	Sequence 28, Appl	528	6	0.7	429	1	US-07-602-824A-2	Sequence 2, Appl
456	6	0.7	381	2	US-08-845-566-3	Sequence 3, Appl	529	6	0.7	429	1	US-07-602-824A-2	Sequence 3, Appl
457	6	0.7	381	3	US-08-467-948A-28	Sequence 28, Appl	530	6	0.7	429	1	US-07-983-451-2	Sequence 2, Appl
458	6	0.7	381	3	US-08-852-824-18	Sequence 12, Appl	531	6	0.7	429	1	US-08-261-577-7	Sequence 7, Appl
459	6	0.7	381	3	US-08-467-947A-28	Sequence 12, Appl	532	6	0.7	429	1	US-08-261-577-7	Sequence 9, Appl
460	6	0.7	381	5	PCT-US96-10618-4	Sequence 2, Appl	533	6	0.7	429	1	US-09-070-356-4	Sequence 4, Appl
461	6	0.7	381	5	US-09-542-733-2	Sequence 2, Appl	534	6	0.7	438	4	US-09-134-001C-3139	Sequence 3139, Ap
462	6	0.7	382	4	US-08-262-477-2	Sequence 4, Appl	535	6	0.7	438	5	PCT-US95-05922A-2	Patent No. 5229279
463	6	0.7	382	4	US-08-196-989B-4	Sequence 4, Appl	536	6	0.7	439	3	US-08-993-359-24	Sequence 24, Appl
464	6	0.7	383	1			537	6	0.7				
465	6	0.7					538	6	0.7				



539	6	0.7	439	3	US-09-221-654-2	Sequence 2, Appli	612	6	0.7	522	4	US-08-894-818B-3	Sequence 3, Appli
540	6	0.7	439	3	US-08-989-358A-2	Sequence 2, Appli	613	6	0.7	522	4	US-09-445-472-4	Sequence 114, App
541	6	0.7	441	2	US-08-713-815A-3	Sequence 3, Appli	614	6	0.7	523	2	US-08-997-080-114	Sequence 19, Appl
542	6	0.7	442	3	US-08-993-359-26	Sequence 26, Appl	615	6	0.7	523	2	US-08-482-728A-19	Sequence 2, Appli
543	6	0.7	445	3	US-08-859-167-2	Sequence 2, Appli	616	6	0.7	523	2	US-08-923-772-2	Sequence 114, App
544	6	0.7	445	3	US-09-109-273-2	Sequence 2, Appli	617	6	0.7	523	2	US-08-997-362-114	Sequence 4, Appli
545	6	0.7	445	3	US-09-276-993-2	Sequence 2, Appli	618	6	0.7	523	4	US-09-028-366-4	Sequence 114, App
546	6	0.7	448	2	US-08-884-072-1	Sequence 1, Appli	619	6	0.7	523	4	US-09-095-855-114	Sequence 114, App
547	6	0.7	448	2	US-09-212-168-1	Sequence 1, Appli	620	6	0.7	523	4	US-09-324-542-114	Sequence 114, App
548	6	0.7	452	1	US-08-191-337-3	Sequence 3, Appli	621	6	0.7	523	4	US-09-385-287-2	Sequence 2, Appli
549	6	0.7	452	4	US-09-323-872A-15	Sequence 15, Appl	622	6	0.7	523	4	US-09-205-426-114	Sequence 114, App
550	6	0.7	454	4	US-09-518-046-2	Sequence 2, Appli	623	6	0.7	524	3	US-08-957-302A-2	Sequence 2, Appli
551	6	0.7	455	2	US-08-870-827-3	Sequence 3, Appli	624	6	0.7	524	4	US-09-299-662-1	Sequence 1, Appli
552	6	0.7	455	4	US-09-308-003-13	Sequence 13, Appli	625	6	0.7	524	4	US-09-542-403-2	Sequence 2, Appli
553	6	0.7	455	4	US-09-317-179-3	Sequence 3, Appli	626	6	0.7	525	1	US-08-681-129-2	Sequence 2, Appli
554	6	0.7	456	1	US-08-021-601-6	Sequence 6, Appli	627	6	0.7	526	2	US-08-541-033A-2	Sequence 2, Appli
555	6	0.7	456	1	US-08-082-849B-6	Sequence 6, Appli	628	6	0.7	526	2	US-08-828-451-2	Sequence 2, Appli
556	6	0.7	456	5	PCT-US94-01624-6	Sequence 6, Appli	629	6	0.7	530	4	US-08-793-044-3	Sequence 3, Appli
557	6	0.7	457	5	US-09-457-046B-61	Sequence 61, Appl	630	6	0.7	534	3	US-08-875-223-8	Sequence 8, Appli
558	6	0.7	458	4	US-09-305-984-20	Sequence 20, Appl	631	6	0.7	538	4	US-09-370-368-9	Sequence 9, Appli
559	6	0.7	459	4	US-09-073-541A-20	Sequence 20, Appl	632	6	0.7	539	4	US-09-309-572-12	Sequence 12, Appl
560	6	0.7	462	1	US-08-417-330A-16	Sequence 16, Appl	633	6	0.7	539	3	US-08-941-445A-21	Sequence 21, Appl
561	6	0.7	465	2	US-08-878-989-18	Sequence 18, Appl	634	6	0.7	540	4	US-09-011-074-4	Sequence 4, Appli
562	6	0.7	465	2	US-08-860-150-7	Sequence 7, Appli	635	6	-0.7	540	4	US-08-461-722-4	Sequence 4, Appli
563	6	0.7	465	3	US-09-338-132-7	Sequence 7, Appli	636	6	0.7	540	4	US-08-336-251-4	Sequence 4, Appli
564	6	0.7	465	4	US-09-272-796-18	Sequence 18, Appl	637	6	0.7	540	5	PCT-US94-06362-4	Sequence 4, Appli
565	6	0.7	465	4	US-09-134-001C-3856	Sequence 3856, Ap	638	6	0.7	541	2	US-08-997-080-160	Sequence 160, App
566	6	0.7	472	1	US-08-021-601-8	Sequence 8, Appli	639	6	0.7	541	2	US-08-997-362-160	Sequence 160, App
567	6	0.7	472	1	US-08-082-849B-8	Sequence 8, Appli	640	6	0.7	541	4	US-09-095-855-160	Sequence 160, App
568	6	0.7	472	5	PCT-US94-01624-8	Sequence 8, Appli	641	6	0.7	541	4	US-09-324-542-160	Sequence 160, App
569	6	0.7	479	4	US-09-177-349-3	Sequence 3, Appli	642	6	0.7	541	4	US-09-324-542-160	Sequence 160, App
570	6	0.7	480	4	US-08-795-088A-2	Sequence 2, Appli	643	6	0.7	542	4	US-09-205-426-160	Sequence 160, App
571	6	0.7	480	4	US-09-069-023-34	Sequence 34, Appl	644	6	0.7	545	2	US-08-467-822-30	Sequence 30, Appl
572	6	0.7	482	1	US-07-792-885A-1	Sequence 1, Appli	645	6	0.7	545	4	US-08-432-697-30	Sequence 30, Appl
573	6	0.7	482	1	US-08-142-439A-7	Sequence 7, Appli	646	6	0.7	545	4	US-08-466-248-30	Sequence 30, Appl
574	6	0.7	482	2	US-08-869-477-7	Sequence 7, Appli	647	6	0.7	546	3	US-08-470-260-6	Sequence 6, Appli
575	6	0.7	482	4	US-09-155-183-4	Sequence 4, Appli	648	6	0.7	546	3	US-08-471-491-6	Sequence 6, Appli
576	6	0.7	488	4	US-09-039-053-2	Sequence 2, Appli	649	6	0.7	546	4	US-08-466-662-6	Sequence 6, Appli
577	6	0.7	488	4	US-08-939-309-10	Sequence 10, Appl	650	6	0.7	546	4	US-08-256-847C-1	Sequence 1, Appli
578	6	0.7	493	4	US-08-999-774A-12	Sequence 12, Appl	651	6	0.7	546	4	US-08-256-847C-7	Sequence 7, Appli
579	6	0.7	501	2	US-08-980-060-6	Sequence 6, Appli	652	6	0.7	547	4	US-09-312-183A-2	Sequence 2, Appli
580	6	0.7	501	4	US-09-307-185-6	Sequence 6, Appli	653	6	0.7	550	4	US-09-330-245A-2	Sequence 2, Appli
581	6	0.7	503	4	US-08-740-223A-11	Sequence 11, Appl	654	6	0.7	553	2	US-08-387-942C-2	Sequence 2, Appli
582	6	0.7	503	4	US-09-709-188-11	Sequence 11, Appl	655	6	0.7	557	4	US-09-518-386B-1	Sequence 1, Appli
583	6	0.7	506	4	US-09-299-662-4	Sequence 4, Appli	656	6	0.7	557	4	US-09-518-386B-3	Sequence 3, Appli
584	6	0.7	508	1	US-08-021-601-10	Sequence 10, Appl	657	6	0.7	567	4	US-09-347-878-42	Sequence 42, Appl
585	6	0.7	508	1	US-08-082-849B-10	Sequence 10, Appl	658	6	0.7	568	1	US-08-262-424-7	Sequence 7, Appli
586	6	0.7	508	5	PCT-US94-01624-10	Sequence 10, Appl	659	6	0.7	568	2	US-08-493-197-7	Sequence 7, Appli
587	6	0.7	509	1	US-08-149-105-14	Sequence 14, Appl	660	6	0.7	568	2	US-08-717-567A-4	Sequence 4, Appli
588	6	0.7	509	1	US-08-317-847-14	Sequence 14, Appl	661	6	0.7	568	3	US-08-883-610A-4	Sequence 4, Appli
589	6	0.7	509	2	US-08-665-926-8	Sequence 8, Appli	662	6	0.7	568	4	US-09-238-373-2	Sequence 2, Appli
590	6	0.7	509	2	US-08-481-337A-4	Sequence 4, Appli	663	6	0.7	568	4	US-08-936-094A-4	Sequence 4, Appli
591	6	0.7	509	4	US-09-382-256-4	Sequence 4, Appli	664	6	0.7	568	4	US-08-939-309-4	Sequence 4, Appli
592	6	0.7	509	4	US-08-740-223A-10	Sequence 10, Appl	665	6	0.7	568	5	PCT-US95-07844-7	Sequence 7, Appli
593	6	0.7	509	4	US-09-395-115-4	Sequence 4, Appli	666	6	0.7	573	1	US-08-200-512-2	Sequence 2, Appli
594	6	0.7	509	4	US-08-123-934A-6	Sequence 6, Appli	667	6	0.7	577	4	US-09-261-855-2	Sequence 4, Appli
595	6	0.7	509	4	US-08-436-265-4	Sequence 4, Appli	668	6	0.7	579	4	US-09-171-710-6	Sequence 6, Appli
596	6	0.7	509	4	US-09-679-187-4	Sequence 4, Appli	669	6	0.7	583	3	US-08-961-862-2	Sequence 3, Appli
597	6	0.7	509	4	US-09-202-491-2	Sequence 2, Appli	670	6	0.7	585	4	US-09-173-300-7	Sequence 7, Appli
598	6	0.7	509	4	US-09-202-491-3	Sequence 3, Appli	671	6	0.7	589	1	US-07-668-648-6	Sequence 6, Appli
599	6	0.7	509	5	US-09-709-188-10	Sequence 10, Appl	672	6	0.7	589	2	US-08-724-394A-6	Sequence 6, Appli
600	6	0.7	509	5	PCT-US94-10080-6	Sequence 6, Appli	673	6	0.7	589	2	US-08-429-998-6	Sequence 6, Appli
601	6	0.7	509	5	PCT-US94-11328A-2	Sequence 2, Appli	674	6	0.7	589	2	US-08-431-333-6	Sequence 6, Appli
602	6	0.7	512	1	US-07-623-953-3	Sequence 3, Appli	675	6	0.7	589	4	US-08-991-862-2	Sequence 2, Appli
603	6	0.7	512	2	US-08-541-033A-4	Sequence 4, Appli	676	6	0.7	589	5	PCT-US91-02321-6	Sequence 6, Appli
604	6	0.7	512	2	US-08-828-451-4	Sequence 4, Appli	677	6	0.7	590	2	US-08-736-317-12	Sequence 12, Appl
605	6	0.7	513	3	US-08-948-564-2	Sequence 2, Appli	678	6	0.7	590	2	US-09-134-001C-4515	Sequence 4515, Ap
606	6	0.7	515	3	US-09-146-283-2	Sequence 2, Appli	679	6	0.7	592	2	US-08-599-171A-30	Sequence 30, Appl
607	6	0.7	515	3	US-08-579-823A-2	Sequence 2, Appli	680	6	0.7	592	2	US-08-646-590B-30	Sequence 30, Appl
608	6	0.7	515	3	US-09-344-195-2	Sequence 2, Appli	681	6	0.7	592	3	US-09-069-226-30	Sequence 30, Appl
609	6	0.7	519	4	US-09-312-183A-3	Sequence 3, Appli	682	6	0.7	592	4	US-09-412-184-30	Sequence 30, Appl
610	6	0.7	521	4	US-09-071-035-64	Sequence 64, Appl	683	6	0.7	594	4	US-08-987-131-2	Sequence 2, Appli
611	6	0.7	521	4	US-09-071-035-64	Sequence 64, Appl	684	6	0.7	594	4	US-08-987-131-2	Sequence 2, Appli

685	6	0.7	596	4	US-09-134-001C-3363	Sequence 3363, Ap	758	6	0.7	682	2	US-08-911-445-3	Sequence 3, Appl1
686	6	0.7	597	6	5171850-2	Patent No. 5171850	759	6	0.7	682	4	US-09-182-983-3	Sequence 3, Appl1
687	6	0.7	599	1	US-07-879-617A-10	Sequence 10, Appl	760	6	0.7	684	4	US-09-061-764A-15	Sequence 15, Appl
688	6	0.7	599	1	US-08-295-814E-11	Sequence 11, Appl	761	6	0.7	686	4	US-09-071-035-46A	Sequence 46A, App
689	6	0.7	599	1	US-08-240-783B-4	Sequence 4, Appl1	762	6	0.7	688	4	US-08-235-836C-68	Sequence 68, Appl
690	6	0.7	599	1	US-08-753-985-10	Sequence 10, Appl	763	6	0.7	693	4	US-08-935-836C-72	Sequence 72, Appl
691	6	0.7	599	3	US-09-084-813-4	Sequence 4, Appl1	764	6	0.7	693	4	US-08-960-048-11	Sequence 11, Appl
692	6	0.7	599	3	US-09-343-361-11	Sequence 11, Appl	765	6	0.7	698	4	US-08-941-445A-11	Sequence 11, Appl
693	6	0.7	599	3	US-09-084-813-4	Sequence 4, Appl1	766	6	0.7	703	4	US-08-910-925-4	Sequence 4, Appl1
694	6	0.7	599	3	PCT-US92-09662-4	Sequence 12, Appl	767	6	0.7	715	3	US-08-425-843-7	Sequence 7, Appl1
695	6	0.7	600	4	US-09-212-971-12	Sequence 12, Appl	768	6	0.7	715	3	US-08-425-843-7	Sequence 7, Appl1
696	6	0.7	600	4	US-08-800-929A-12	Sequence 12, Appl	769	6	0.7	725	3	US-08-425-843-221	Sequence 321, App
697	6	0.7	600	4	US-09-617-053A-12	Sequence 12, Appl	770	6	0.7	729	4	US-09-625-188-20	Sequence 20, Appl
698	6	0.7	604	3	US-08-511-485-6	Sequence 6, Appl1	771	6	0.7	745	1	US-08-453-472-5	Sequence 5, Appl1
699	6	0.7	604	3	US-09-212-971-6	Sequence 6, Appl1	772	6	0.7	745	1	US-08-038-948-9	Sequence 9, Appl1
700	6	0.7	604	3	US-08-800-929A-6	Sequence 6, Appl1	773	6	0.7	745	1	US-08-038-948-9	Sequence 9, Appl1
701	6	0.7	604	4	US-08-569-749-4	Sequence 4, Appl1	774	6	0.7	745	1	US-08-445-952-5	Sequence 5, Appl1
702	6	0.7	604	4	US-09-617-053A-6	Sequence 6, Appl1	775	6	0.7	745	2	US-08-484-993B-43	Sequence 43, Appl
703	6	0.7	607	5	PCT-US96-12860-4	Sequence 4, Appl1	776	6	0.7	745	2	US-08-484-993B-43	Sequence 43, Appl
704	6	0.7	607	4	US-08-969-683A-59	Sequence 59, Appl	777	6	0.7	745	2	US-08-862-903-5	Sequence 5, Appl1
705	6	0.7	609	4	US-09-134-001C-2994	Sequence 2994, Ap	778	6	0.7	745	2	US-08-862-903-5	Sequence 5, Appl1
706	6	0.7	609	4	US-09-115-475-16	Sequence 16, Appl	779	6	0.7	745	2	US-08-484-158B-43	Sequence 43, Appl
707	6	0.7	609	4	US-09-115-475-19	Sequence 19, Appl	780	6	0.7	745	2	US-08-484-158B-43	Sequence 43, Appl
708	6	0.7	609	4	US-09-115-475-22	Sequence 22, Appl	781	6	0.7	745	2	US-08-480-150A-43	Sequence 43, Appl
709	6	0.7	609	4	US-09-115-475-25	Sequence 25, Appl	782	6	0.7	745	2	US-08-480-150A-43	Sequence 43, Appl
710	6	0.7	609	4	US-09-115-475-31	Sequence 31, Appl	783	6	0.7	747	3	US-08-149-223A-43	Sequence 43, Appl
711	6	0.7	612	3	US-09-212-971-14	Sequence 14, Appl	784	6	0.7	747	3	PCT-US95-10509-2	Sequence 2, Appl1
712	6	0.7	612	3	US-08-800-929A-14	Sequence 14, Appl	785	6	0.7	747	3	US-08-816-693A-51	Sequence 51, Appl
713	6	0.7	612	4	US-08-569-749-14	Sequence 14, Appl	786	6	0.7	747	4	US-08-885-291-51	Sequence 51, Appl
714	6	0.7	612	4	US-09-617-053A-14	Sequence 14, Appl	787	6	0.7	749	4	US-09-496-672-51	Sequence 51, Appl
715	6	0.7	612	4	US-09-617-053A-14	Sequence 14, Appl	788	6	0.7	754	4	US-09-562-737-84	Sequence 84, Appl
716	6	0.7	612	5	PCT-US96-12860-14	Sequence 14, Appl	789	6	0.7	754	4	US-09-521-780-2	Sequence 2, Appl1
717	6	0.7	615	2	US-08-911-445-2	Sequence 2, Appl1	790	6	0.7	766	4	US-09-521-780-3	Sequence 3, Appl1
718	6	0.7	615	2	US-09-182-983-2	Sequence 2, Appl1	791	6	0.7	766	4	US-09-521-780-3	Sequence 3, Appl1
719	6	0.7	616	2	US-08-545-745-2	Sequence 2, Appl1	792	6	0.7	772	3	US-08-021-601-2	Sequence 12, Appl
720	6	0.7	618	3	US-09-212-971-8	Sequence 8, Appl1	793	6	0.7	776	1	US-08-082-849B-2	Sequence 2, Appl1
721	6	0.7	618	3	US-08-511-485-8	Sequence 8, Appl1	794	6	0.7	776	5	PCT-US94-01624-2	Sequence 2, Appl1
722	6	0.7	618	4	US-08-800-929A-8	Sequence 8, Appl1	795	6	0.7	788	2	US-08-907-166-6	Sequence 2, Appl1
723	6	0.7	618	4	US-09-569-749-2	Sequence 2, Appl1	796	6	0.7	797	4	US-08-945-983-2	Sequence 4, Appl1
724	6	0.7	618	4	US-09-617-053A-8	Sequence 8, Appl1	797	6	0.7	806	4	US-09-180-439-6	Sequence 6, Appl1
725	6	0.7	618	4	US-09-069-023-29	Sequence 29, Appl	798	6	0.7	809	4	US-08-945-983-2	Sequence 2, Appl1
726	6	0.7	618	5	US-09-134-001C-4039	Sequence 4039, Ap	799	6	0.7	812	1	US-09-105-537-24	Sequence 24, Appl
727	6	0.7	618	5	PCT-US96-12860-2	Sequence 2, Appl1	800	6	0.7	812	1	US-08-248-629A-1	Sequence 1, Appl1
728	6	0.7	621	1	US-08-934-361-6	Sequence 4248, Ap	801	6	0.7	812	1	US-08-451-932-1	Sequence 1, Appl1
729	6	0.7	621	5	PCT-US93-01959-6	Sequence 6, Appl1	802	6	0.7	812	1	US-08-452-260-1	Sequence 1, Appl1
730	6	0.7	621	5	US-09-343-361-6	Sequence 6, Appl1	803	6	0.7	812	2	US-08-326-785-1	Sequence 1, Appl1
731	6	0.7	623	4	US-09-104-068-4	Sequence 4, Appl1	804	6	0.7	812	2	US-08-612-788-1	Sequence 1, Appl1
732	6	0.7	623	3	US-08-581-148C-18	Sequence 18, Appl	805	6	0.7	812	2	US-08-605-598B-1	Sequence 1, Appl1
733	6	0.7	625	4	US-08-759-436-3	Sequence 4, Appl1	806	6	0.7	812	2	US-08-429-743-1	Sequence 1, Appl1
734	6	0.7	625	4	US-08-897-443-1	Sequence 4, Appl1	807	6	0.7	812	3	US-08-866-735-1	Sequence 1, Appl1
735	6	0.7	637	4	US-09-104-068-2	Sequence 2, Appl1	808	6	0.7	819	4	US-09-066-028-1	Sequence 1, Appl1
736	6	0.7	637	4	US-08-759-436-5	Sequence 5, Appl1	809	6	0.7	819	4	PCT-US95-05107-1	Sequence 1, Appl1
737	6	0.7	638	2	US-09-026-343-2	Sequence 2, Appl1	810	6	0.7	819	4	US-09-651-656-15	Sequence 15, Appl
738	6	0.7	640	4	US-08-897-443-1	Sequence 1, Appl1	811	6	0.7	821	1	US-09-650-855-15	Sequence 15, Appl
739	6	0.7	640	4	US-09-362-871-2	Sequence 2, Appl1	812	6	0.7	821	1	US-07-935-311A-4	Sequence 4, Appl1
740	6	0.7	643	2	US-08-797-358B-3	Sequence 3, Appl1	813	6	0.7	821	5	US-08-368-0799-4	Sequence 5, Appl1
741	6	0.7	647	2	US-08-305-764C-56	Sequence 56, Appl	814	6	0.7	830	4	US-08-368-0799-4	Sequence 4, Appl1
742	6	0.7	647	2	US-08-844-056-2	Sequence 2, Appl1	815	6	0.7	837	4	PCT-US93-07996-4	Sequence 36, Appl
743	6	0.7	647	2	US-09-300-909-17	Sequence 17, Appl	816	6	0.7	837	4	US-09-562-737-36	Sequence 3, Appl1
744	6	0.7	649	4	US-09-192-909-2	Sequence 2, Appl1	817	6	0.7	842	2	US-09-122-126B-2	Sequence 2, Appl1
745	6	0.7	650	3	US-08-362-525-2	Sequence 2, Appl1	818	6	0.7	842	4	US-08-928-692-18	Sequence 18, Appl
746	6	0.7	652	4	US-08-987-123-2	Sequence 2, Appl1	819	6	0.7	843	3	US-09-339-972-18	Sequence 18, Appl
747	6	0.7	654	4	US-09-061-764A-2	Sequence 2, Appl1	820	6	0.7	845	2	US-09-215-966-22	Sequence 22, Appl
748	6	0.7	654	4	US-08-894-818B-35	Sequence 35, Appl	821	6	0.7	845	2	US-08-416-950-11	Sequence 11, Appl
749	6	0.7	654	4	US-09-445-472-16	Sequence 16, Appl	822	6	0.7	852	1	US-08-469-830-11	Sequence 11, Appl
750	6	0.7	659	4	US-08-894-818B-5	Sequence 5, Appl1	823	6	0.7	852	1	US-08-190-802A-59	Sequence 59, Appl
751	6	0.7	659	4	US-09-228-986-75	Sequence 75, Appl1	824	6	0.7	852	4	US-08-190-802A-59	Sequence 59, Appl
752	6	0.7	663	3	US-08-894-818B-5	Sequence 5, Appl1	825	6	0.7	852	4	US-08-477-346-59	Sequence 59, Appl
753	6	0.7	675	3	US-08-776-265-5	Sequence 5, Appl1	826	6	0.7	852	4	US-08-473-346-67	Sequence 67, Appl
754	6	0.7	676	4	US-08-947-965-76	Sequence 76, Appl1	827	6	0.7	852	4	US-08-473-346-67	Sequence 67, Appl
755	6	0.7	676	4	US-08-760-615-2	Sequence 2, Appl1	828	6	0.7	852	4	US-08-487-072A-59	Sequence 59, Appl
756	6	0.7	679	4	US-09-693-147-6	Sequence 6, Appl1	829	6	0.7	861	3	US-08-487-072A-59	Sequence 59, Appl
757	6	0.7	679	4	US-09-177-650-4	Sequence 15, Appl	830	6	0.7	863	3	US-09-022-875-2	Sequence 2, Appl1
	6	0.7	679	4	US-08-913-942-15	Sequence 15, Appl							
	6	0.7	679	4	US-09-268-347-26	Sequence 26, Appl							

831	6	0.7	876	2	US-08-633-476-2	Sequence 2, Appl	904	6	0.7	1130	2	US-08-125-077-2	Sequence 2, Appl
832	6	0.7	876	2	US-08-436-664-20	Sequence 30, Appl	905	6	0.7	1130	6	5444158-2	Patent No. 5444158
833	6	0.7	876	2	US-08-436-664-32	Sequence 32, Appl	906	6	0.7	1150	4	US-09-238-303-9	Sequence 9, Appl
834	6	0.7	876	2	US-08-436-664-34	Sequence 34, Appl	907	6	0.7	1155	1	US-08-286-889-46	Sequence 46, Appl
835	6	0.7	876	3	US-09-135-642-20	Sequence 20, Appl	908	6	0.7	1155	1	US-08-485-618-46	Sequence 46, Appl
836	6	0.7	876	3	US-09-135-642-32	Sequence 32, Appl	909	6	0.7	1155	1	US-08-362-652-46	Sequence 46, Appl
837	6	0.7	876	3	US-09-135-642-34	Sequence 34, Appl	910	6	0.7	1155	2	US-08-605-672-46	Sequence 46, Appl
838	6	0.7	876	3	US-08-394-232A-20	Sequence 20, Appl	911	6	0.7	1155	2	US-08-482-293A-46	Sequence 46, Appl
839	6	0.7	876	3	US-08-394-232A-32	Sequence 32, Appl	912	6	0.7	1155	2	US-08-482-293A-46	Sequence 46, Appl
840	6	0.7	876	3	US-08-394-232A-34	Sequence 34, Appl	913	6	0.7	1155	2	US-09-193-043-46	Sequence 46, Appl
841	6	0.7	876	4	US-09-517-871-2	Sequence 2, Appl	914	6	0.7	1155	4	US-09-688-307A-46	Sequence 46, Appl
842	6	0.7	876	4	US-09-517-871-2	Sequence 2, Appl	915	6	0.7	1156	4	US-09-002-285-72	Sequence 72, Appl
843	6	0.7	876	5	PCT-US95-04080-20	Sequence 20, Appl	916	6	0.7	1157	2	US-08-532-547-5	Sequence 5, Appl
844	6	0.7	876	5	PCT-US95-04080-32	Sequence 32, Appl	917	6	0.7	1157	2	US-08-379-656B-5	Sequence 5, Appl
845	6	0.7	876	5	PCT-US95-04080-34	Sequence 34, Appl	918	6	0.7	1157	3	US-08-455-838-5	Sequence 5, Appl
846	6	0.7	877	1	US-08-208-036-14	Sequence 14, Appl	919	6	0.7	1157	4	US-09-019-809-5	Sequence 5, Appl
847	6	0.7	877	1	US-08-428-823-14	Sequence 14, Appl	920	6	0.7	1157	4	US-09-471-177-5	Sequence 5, Appl
848	6	0.7	906	2	US-08-609-230A-9	Sequence 9, Appl	921	6	0.7	1161	1	US-08-485-618-53	Sequence 53, Appl
849	6	0.7	907	3	US-08-990-140-4	Sequence 4, Appl	922	6	0.7	1161	1	US-08-362-652-53	Sequence 53, Appl
850	6	0.7	907	4	US-09-546-238-4	Sequence 4, Appl	923	6	0.7	1161	2	US-08-605-672-53	Sequence 53, Appl
851	6	0.7	913	1	US-08-487-890A-5	Sequence 5, Appl	924	6	0.7	1161	2	US-08-482-293A-53	Sequence 53, Appl
852	6	0.7	913	1	US-08-445-640-4	Sequence 4, Appl	925	6	0.7	1161	2	US-08-943-363-53	Sequence 53, Appl
853	6	0.7	913	2	US-08-478-435-5	Sequence 5, Appl	926	6	0.7	1161	4	US-09-193-043-53	Sequence 53, Appl
854	6	0.7	913	2	US-08-337-483-5	Sequence 5, Appl	927	6	0.7	1161	4	US-09-588-307A-53	Sequence 53, Appl
855	6	0.7	913	2	US-08-478-373-5	Sequence 5, Appl	928	6	0.7	1169	2	US-08-542-921-2	Sequence 2, Appl
856	6	0.7	913	3	US-08-170-558-4	Sequence 4, Appl	929	6	0.7	1169	2	US-08-880-685-2	Sequence 2, Appl
857	6	0.7	913	3	US-08-474-671-5	Sequence 5, Appl	930	6	0.7	1169	2	US-08-880-685-2	Sequence 2, Appl
858	6	0.7	913	3	US-08-483-577A-5	Sequence 5, Appl	931	6	0.7	1182	4	US-09-041-886-21	Sequence 21, Appl
859	6	0.7	913	3	US-08-447-314-4	Sequence 4, Appl	932	6	0.7	1185	4	US-09-134-001C-5276	Sequence 5276, Ap
860	6	0.7	913	3	US-08-445-461-4	Sequence 4, Appl	933	6	0.7	1207	1	US-07-951-715A-7	Sequence 7, Appl
861	6	0.7	913	4	US-08-897-438-5	Sequence 5, Appl	934	6	0.7	1207	3	US-08-459-448A-7	Sequence 7, Appl
862	6	0.7	913	4	US-08-637-654-5	Sequence 5, Appl	935	6	0.7	1207	3	US-08-459-595A-7	Sequence 7, Appl
863	6	0.7	913	4	US-08-649-518-5	Sequence 5, Appl	936	6	0.7	1207	3	US-08-459-504B-7	Sequence 7, Appl
864	6	0.7	917	4	US-09-049-698-41	Sequence 41, Appl	937	6	0.7	1207	3	US-08-459-444-7	Sequence 7, Appl
865	6	0.7	919	1	US-08-336-343A-2	Sequence 2, Appl	938	6	0.7	1207	3	US-09-053-549-8	Sequence 8, Appl
866	6	0.7	936	5	PCT-US93-05944-2	Sequence 2, Appl	939	6	0.7	1207	4	US-09-547-422-7	Sequence 7, Appl
867	6	0.7	940	4	US-08-810-712-7	Sequence 7, Appl	940	6	0.7	1213	1	US-08-188-582-20	Sequence 20, Appl
868	6	0.7	947	4	US-09-228-986-73	Sequence 73, Appl	941	6	0.7	1213	1	US-08-646-715-20	Sequence 20, Appl
869	6	0.7	954	3	US-09-057-969-2	Sequence 2, Appl	942	6	0.7	1227	1	US-08-448-170-8	Sequence 8, Appl
870	6	0.7	956	2	US-08-897-443-3	Sequence 3, Appl	943	6	0.7	1227	3	US-09-053-549-2	Sequence 2, Appl
871	6	0.7	956	4	US-09-134-001C-4452	Sequence 4452, Ap	944	6	0.7	1227	4	US-08-961-803-9	Sequence 9, Appl
872	6	0.7	968	4	US-09-180-439-3	Sequence 3, Appl	945	6	0.7	1262	4	US-09-357-251-33	Sequence 33, Appl
873	6	0.7	968	4	US-09-180-439-4	Sequence 4, Appl	946	6	0.7	1266	1	US-08-468-557-4	Sequence 4, Appl
874	6	0.7	968	4	US-09-228-986-76	Sequence 76, Appl	947	6	0.7	1266	1	US-09-357-251-32	Sequence 32, Appl
875	6	0.7	979	4	US-09-323-872A-55	Sequence 55, Appl	948	6	0.7	1306	4	US-08-999-774A-13	Sequence 13, Appl
876	6	0.7	983	4	US-09-134-001C-3814	Sequence 3814, Ap	949	6	0.7	1333	4	US-09-347-878-20	Sequence 20, Appl
877	6	0.7	984	1	US-08-257-073-3	Sequence 3, Appl	950	6	0.7	1365	4	US-09-376-330-18	Sequence 18, Appl
878	6	0.7	984	2	US-08-184-009-120	Sequence 120, App	951	6	0.7	1365	6	5194600-4	Patent No. 5194600
879	6	0.7	984	2	US-08-458-356-120	Sequence 120, App	952	6	0.7	1404	2	US-08-400-159-2	Sequence 2, Appl
880	6	0.7	984	4	US-08-460-736-120	Sequence 120, App	953	6	0.7	1404	3	US-08-611-729A-2	Sequence 2, Appl
881	6	0.7	988	2	US-08-286-819A-19	Sequence 19, App	954	6	0.7	1422	4	US-08-469-260A-85	Sequence 85, Appl
882	6	0.7	988	3	US-08-980-357-19	Sequence 19, Appl	955	6	0.7	1430	3	US-09-008-172-2	Sequence 2, Appl
883	6	0.7	994	3	US-08-699-103B-3	Sequence 3, Appl	956	6	0.7	1430	4	US-09-210-361-6	Sequence 6, Appl
884	6	0.7	994	4	US-09-229-059-3	Sequence 3, Appl	957	6	0.7	1430	4	US-09-740-274-6	Sequence 6, Appl
885	6	0.7	1004	4	US-09-268-347-30	Sequence 30, Appl	958	6	0.7	1528	1	US-08-463-092B-6	Sequence 6, Appl
886	6	0.7	1009	2	US-08-357-642A-1	Sequence 1, Appl	959	6	0.7	1528	2	US-08-462-109A-6	Sequence 6, Appl
887	6	0.7	1009	2	US-08-460-626-1	Sequence 1, Appl	960	6	0.7	1528	2	US-08-460-907B-6	Sequence 6, Appl
888	6	0.7	1016	4	US-09-180-439-8	Sequence 8, Appl	961	6	0.7	1528	3	US-08-463-179A-6	Sequence 6, Appl
889	6	0.7	1024	4	US-09-562-737-48	Sequence 48, Appl	962	6	0.7	1528	3	US-08-461-384B-6	Sequence 6, Appl
890	6	0.7	1031	4	US-08-811-682-15	Sequence 15, Appl	963	6	0.7	1531	1	US-08-461-384B-6	Sequence 6, Appl
891	6	0.7	1044	2	US-08-777-405A-2	Sequence 2, Appl	964	6	0.7	1531	1	US-08-141-893-3	Sequence 3, Appl
892	6	0.7	1044	2	US-08-977-871A-2	Sequence 2, Appl	965	6	0.7	1531	1	US-08-463-092B-2	Sequence 2, Appl
893	6	0.7	1044	2	US-09-225-951-2	Sequence 2, Appl	966	6	0.7	1531	2	US-08-462-109A-2	Sequence 2, Appl
894	6	0.7	1048	4	US-09-171-699-10	Sequence 10, Appl	967	6	0.7	1531	2	US-08-462-109A-4	Sequence 4, Appl
895	6	0.7	1048	4	US-08-887-534A-85	Sequence 85, Appl	968	6	0.7	1531	2	US-08-460-907B-2	Sequence 2, Appl
896	6	0.7	1073	4	US-09-206-942-49	Sequence 49, Appl	969	6	0.7	1531	2	US-08-460-907B-4	Sequence 4, Appl
897	6	0.7	1073	4	US-09-206-942-47	Sequence 47, Appl	970	6	0.7	1531	3	US-08-463-179A-2	Sequence 2, Appl
898	6	0.7	1104	4	US-09-268-347-28	Sequence 28, Appl	971	6	0.7	1531	3	US-08-463-179A-2	Sequence 2, Appl
899	6	0.7	1104	4	US-09-268-347-34	Sequence 34, Appl	972	6	0.7	1531	3	US-08-461-384B-2	Sequence 2, Appl
900	6	0.7	1112	4	US-09-353-585-2	Sequence 2, Appl	973	6	0.7	1531	3	US-08-461-384B-4	Sequence 4, Appl
901	6	0.7	1112	4	US-09-353-585-3	Sequence 3, Appl	974	6	0.7	1531	3	US-08-407-207A-2	Sequence 2, Appl
902	6	0.7	1118	4	US-09-379-523-3	Sequence 3, Appl	975	6	0.7	1566	6	US-08-687-956A-23	Sequence 23, Appl
903	6	0.7	1130	2	US-08-460-309-2	Sequence 2, Appl	976	6	0.7	1676	4	US-08-487-283A-2	Sequence 2, Appl



```

: APPLICANT: Wen, Zilong
: TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
: TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESS: Klauber & Jackson
: STREET: 411 Hackensack Avenue
: CITY: Hackensack
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/820,754
: FILING DATE: 19-MAR-1997
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/212,185
: FILING DATE: 11-MAR-1994
: APPLICATION NUMBER: US 07/980,498
: FILING DATE: 23-NOV-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/854,296
: FILING DATE: 19-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO US93/02569
: FILING DATE: 19-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/126,588
: FILING DATE: 24-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Jackson Esq., David A.
: REGISTRATION NUMBER: 26,742
: REFERENCE/DOCKET NUMBER: 600-1-073 CIP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201 487-5800
: TELEFAX: 201 343-1684
: TELEX: 133521
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 851 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-820-754-2

Query Match 1.0%; Score 9; DB 2; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 457 ESTLEPVIE 465
    |||||||
Db 756 ESTLEPVIE 764

RESULT 4
US-08-956-652-2
: Sequence 2, Application US/08956652
: Patent No. 6013475
: GENERAL INFORMATION:
: APPLICANT: Darnell Jr., James E.
: APPLICANT: Schindler, Christian W.
: APPLICANT: Fu, Xian-Yuan
: APPLICANT: Wen, Zilong
: APPLICANT: Zhong, Zhong
: TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
: TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
: NUMBER OF SEQUENCES: 25
```

```

: CORRESPONDENCE ADDRESS:
: ADDRESS: Klauber & Jackson
: STREET: 411 Hackensack Avenue
: CITY: Hackensack
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/956,652
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/212,185
: FILING DATE: 11-MAR-1994
: APPLICATION NUMBER: US 07/980,498
: FILING DATE: 23-NOV-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/854,296
: FILING DATE: 19-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO US93/02569
: FILING DATE: 19-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/126,588
: FILING DATE: 24-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Jackson Esq., David A.
: REGISTRATION NUMBER: 26,742
: REFERENCE/DOCKET NUMBER: 600-1-073 CIP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201 487-5800
: TELEFAX: 201 343-1684
: TELEX: 133521
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 851 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-956-652-2

Query Match 1.0%; Score 9; DB 3; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 457 ESTLEPVIE 465
    |||||||
Db 756 ESTLEPVIE 764

RESULT 5
US-08-956-869-2
: Sequence 2, Application US/08956689
: Patent No. 6030808
: GENERAL INFORMATION:
: APPLICANT: Darnell Jr., James E.
: APPLICANT: Schindler, Christian W.
: APPLICANT: Fu, Xian-Yuan
: APPLICANT: Wen, Zilong
: APPLICANT: Zhong, Zhong
: TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
: TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESS: Klauber & Jackson
: STREET: 411 Hackensack Avenue
: CITY: Hackensack
: STATE: New Jersey
```

COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,869  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/212,185  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/854,296  
FILING DATE: 19-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO US93/02569  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,588  
FILING DATE: 24-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-073 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ. ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 851 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-956-869-2

Query Match 1.0%; Score 9; DB 3; Length 851;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465  
|||||||  
Db 756 ESTLEPVIE 764

RESULT 6  
US-09-012-710-2  
Sequence 2, Application US/09012710  
Patent No. 6087478  
GENERAL INFORMATION:  
APPLICANT: Vinkemeler, Uwe  
APPLICANT: Morelli, Ismail  
APPLICANT: Darnell, Jr., James E.  
APPLICANT: Kuriyan, John  
TITLE OF INVENTION: A CRYSTAL OF THE N-TERMINAL DOMAIN OF A  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESS: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/012,710  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-194  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ. ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 851 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-09-012-710-2

Query Match 1.0%; Score 9; DB 3; Length 851;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465  
|||||||  
Db 756 ESTLEPVIE 764

RESULT 7  
US-08-948-547-2  
Sequence 2, Application US/08948547  
Patent No. 6124118  
GENERAL INFORMATION:  
APPLICANT: Darnell Jr., James E.  
APPLICANT: Schindler, Christian W.  
APPLICANT: Fu, Xian-Yuan  
APPLICANT: Wen, Zilong  
APPLICANT: Zhong, Zhong  
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESS: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/948,547  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/212,185  
FILING DATE: 11-MAR-1994  
APPLICATION NUMBER: US 07/980,498  
FILING DATE: 23-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/854,296  
FILING DATE: 19-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO US93/02569  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,588  
FILING DATE: 24-SEP-1993

```
ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-073 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 851 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-948-547-2

Query Match      1.0%; Score 9; DB 3; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465
   |||||||
Db 756 ESTLEPVIE 764

RESULT 8
US-09-087-465-4
; Sequence 4, Application US/09087465A
; Patent No. 6160092
; GENERAL INFORMATION:
; APPLICANT: Vinkemeier, Uwe
; APPLICANT: Chen, Xiaomlin
; APPLICANT: Darnell Jr., James E
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF THE CORE PORTION OF A STAT AND METHODS OF
; TITLE OF INVENTION: USE
; FILE REFERENCE: 600-1-229
; CURRENT APPLICATION NUMBER: US/09/087,465A
; CURRENT FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 851
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-087-465-4

Query Match      1.0%; Score 9; DB 4; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465
   |||||||
Db 756 ESTLEPVIE 764

RESULT 9
US-09-364-970-2
; Sequence 2, Application US/09364970
; Patent No. 6235873
; GENERAL INFORMATION:
; APPLICANT: Bromberg, Jacqueline
; TITLE OF INVENTION: CONSTITUTIVELY ACTIVE TRANSCRIPTION FACTORS AND THEIR
; TITLE OF INVENTION: USES FOR IDENTIFYING MODULATORS OF ACTIVITY INCLUDING
; FILE REFERENCE: 600-1-252
; CURRENT APPLICATION NUMBER: US/09/364,970
; CURRENT FILING DATE: 1999-07-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 851
; TYPE: PRT
```

```
ORGANISM: Homo sapiens
US-09-364-970-2

Query Match      1.0%; Score 9; DB 4; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465
   |||||||
Db 756 ESTLEPVIE 764

RESULT 10
US-09-556-273-2
; Sequence 2, Application US/09556273
; Patent No. 6312887
; GENERAL INFORMATION:
; APPLICANT: Vinkemeier, Uwe
; APPLICANT: Moarell, Ismail
; APPLICANT: Darnell, Jr., James E.
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF THE N-TERMINAL DOMAIN OF A
; TITLE OF INVENTION: STAT PROTEIN AND METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/556,273
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/012,710
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-194
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 851 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-556-273-2

Query Match      1.0%; Score 9; DB 4; Length 851;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465
   |||||||
Db 756 ESTLEPVIE 764

RESULT 11
US-08-956-653A-2
; Sequence 2, Application US/08956653A
```

Patent No. 6338949  
GENERAL INFORMATION:  
APPLICANT: Darnell Jr., James E.  
APPLICANT: Schindler, Christian W.  
APPLICANT: Fu, Xian-Yuan  
APPLICANT: Wen, Zilong  
APPLICANT: Zhong, Zhong  
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN  
NUMBER OF SEQUENCES: 34  
SEQUENCES AND METHODS OF USE THEREOF  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,653A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/212,185  
FILING DATE: 11-MAR-1994  
APPLICATION NUMBER: US 07/980,498  
FILING DATE: 23-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/854,296  
FILING DATE: 19-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO US93/02569  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,588  
FILING DATE: 24-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION/DOCKET NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-195  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 851 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-956-653A-2

Query Match 1.0%; Score 9; DB 4; Length 851;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465  
DB 756 ESTLEPVIE 764

RESULT 12  
PCT-US95-17025-2  
Sequence 2, Application PC/TUS9517025  
GENERAL INFORMATION:  
APPLICANT: James E. Darnell, Jr.  
APPLICANT: Zilong Wen  
APPLICANT: Curt M. Horvath  
APPLICANT: Zhong Zhong

TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL  
TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/17025  
FILING DATE: 28-DEC-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/369,796  
FILING DATE: 06-JAN-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION/DOCKET NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 851 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-17025-2

Query Match 1.0%; Score 9; DB 5; Length 851;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 ESTLEPVIE 465  
DB 756 ESTLEPVIE 764

RESULT 13  
US-08-276-099A-13  
Sequence 13, Application US/08276099A  
Patent No. 5591825  
GENERAL INFORMATION:  
APPLICANT: McKnight, Steven L  
APPLICANT: Hou, Jinzhao  
TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND  
BINDING ASSAYS  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HORBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/276,099A  
FILING DATE: 15-JUL-1994



CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard Aron  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59451-1/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 852 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-276-099A-13

Query Match 1.0%; Score 9; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ESTLEPVIE 465  
|||||  
DB 756 ESTLEPVIE 764

RESULT 14  
US-08-781-890-13  
Sequence 13, Application US/08781890  
Patent No. 5710266  
GENERAL INFORMATION:  
APPLICANT: McKnight, Steven L  
APPLICANT: Hou, Jinzhao  
TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND  
TITLE OF INVENTION: BINDING ASSAYS  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,890  
FILING DATE: 05-JAN-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/276,099  
FILING DATE: 15-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard Aron  
REGISTRATION NUMBER: 36,627  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 852 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-781-890-13

Query Match 1.0%; Score 9; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 ESTLEPVIE 465  
|||||  
DB 756 ESTLEPVIE 764

RESULT 15  
US-09-242-632A-2  
Sequence 2, Application US/09242632A  
Patent No. 6312696  
GENERAL INFORMATION:  
APPLICANT: Nippon Zeon  
TITLE OF INVENTION: Antigen protein derived from infectious  
TITLE OF INVENTION: Laryngotracheitis virus  
FILE REFERENCE: E868-PCR  
CURRENT APPLICATION NUMBER: US/09/242,632A  
CURRENT FILING DATE: 1999-02-19  
PRIOR APPLICATION NUMBER: JP 8-238580  
PRIOR FILING DATE: 1996-08-21  
NUMBER OF SEQ ID NOS: 28  
SEQ ID NO 2  
LENGTH: 582  
TYPE: PRT  
ORGANISM: Infectious laryngotracheitis virus  
FEATURE:  
OTHER INFORMATION: Amino acid sequence of 132h polypeptide of  
OTHER INFORMATION: Infectious laryngotracheitis virus  
US-09-242-632A-2

Query Match 0.9%; Score 8; DB 4; Length 582;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 213 KAALEDIT 220  
|||||  
DB 277 KAALEDIT 284

RESULT 16  
US-07-960-510-4  
Sequence 4, Application US/07960510  
Patent No. 5705362  
GENERAL INFORMATION:  
APPLICANT: BONEKAMP, ALFONSUSJOHANNES  
APPLICANT: VAN TILBORG E., MARCELIS W.  
TITLE OF INVENTION: MODIFIED SIGNAL SEQUENCES  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave., Suite 5500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/960,510  
FILING DATE: 19921013  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 24615-20034.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763

TELEX: 90-4030  
INFORMATION FOR SEQ. ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-960-510-4

Query Match  
Best Local Similarity 0.8%; Score 7; DB 1; Length 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 KRLYARL 172  
DB 5 KRLYARL 11

RESULT 17  
US-07-960-510-5  
Sequence 5, Application US/07960510  
Patent No. 5705362

GENERAL INFORMATION:  
APPLICANT: BONEKAMP, ALFONSUSJOHANNES  
APPLICANT: VAN TILBORG E., MARCELLIS W.  
TITLE OF INVENTION: MODIFIED SIGNAL SEQUENCES  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESS: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave., Suite 5500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20006-1888

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/960,510  
FILING DATE: 19921013  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 24615-20034.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030

INFORMATION FOR SEQ. ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-960-510-5

Query Match  
Best Local Similarity 0.8%; Score 7; DB 1; Length 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 KRLYARL 172  
DB 5 KRLYARL 11

RESULT 18  
US-07-960-510-6  
Sequence 6, Application US/07960510  
Patent No. 5705362  
GENERAL INFORMATION:

APPLICANT: BONEKAMP, ALFONSUSJOHANNES  
APPLICANT: VAN TILBORG E., MARCELLIS W.  
TITLE OF INVENTION: MODIFIED SIGNAL SEQUENCES  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESS: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave., Suite 5500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20006-1888

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/960,510  
FILING DATE: 19921013  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 24615-20034.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-960-510-6

Query Match  
Best Local Similarity 0.8%; Score 7; DB 1; Length 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 166 KRLYARL 172  
DB 5 KRLYARL 11

RESULT 19  
US-07-960-510-7  
Sequence 7, Application US/07960510  
Patent No. 5705362  
GENERAL INFORMATION:

APPLICANT: BONEKAMP, ALFONSUSJOHANNES  
APPLICANT: VAN TILBORG E., MARCELLIS W.  
TITLE OF INVENTION: MODIFIED SIGNAL SEQUENCES  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESS: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave., Suite 5500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20006-1888

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/960,510  
FILING DATE: 19921013  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959

```
REFERENCE/DOCKET NUMBER: 24615-20034.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-960-510-7

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 166 KRLYARL 172
Db 5 KRLYARL 11

RESULT 20
5187089-8
; Patent No. 5187089
; APPLICANT: SCOTT, RANDY W.; GOLINI, FRED; MCGROGAN, MICHAEL
; TITLE OF INVENTION: PROTEASE NEXIN-I VARIANTS WHICH INHIBIT
; ELASTASE
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/542,484
; FILING DATE: 21-JUN-1990
; SEQ ID NO: 8:
; LENGTH: 34
5187089-8

Query Match
Best Local Similarity 100.0%; Score 7; DB 6; Length 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 117 PLSTLEL 123
Db 5 PLSTLEL 11

RESULT 21
5457090-5
; Patent No. 5457090
; APPLICANT: SCOTT, RANDY W.; GOLINI, FRED; MCGROGAN, MICHAEL
; TITLE OF INVENTION: PROTEASE NEXIN-I VARIANTS
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/924,294
; FILING DATE: 03-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 342,484
; FILING DATE: 21-JUN-1990
; SEQ ID NO: 5:
; LENGTH: 34
5457090-5

Query Match
Best Local Similarity 100.0%; Score 7; DB 6; Length 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 117 PLSTLEL 123
Db 5 PLSTLEL 11

RESULT 22
5495001-16
; Patent No. 5495001
```

```
APPLICANT: MCGROGAN, MICHAEL P.; SCOTT, RANDY W.; BAKER,
; JOFFRE B.; SIMONSEN, CHRISTIAN C.
; TITLE OF INVENTION: RECOMBINANT PURIFIED PROTEASE NEXIN
; NUMBER OF SEQUENCES: 16
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/288,596
; FILING DATE: 10-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 111,111
; FILING DATE: 24-AUG-1993
; APPLICATION NUMBER: 25,450
; FILING DATE: 13-MAR-1987
; APPLICATION NUMBER: 871,501
; FILING DATE: 06-JUN-1986
; APPLICATION NUMBER: 870,232
; FILING DATE: 03-JUN-1986
; SEQ ID NO: 16:
; LENGTH: 34
5495001-16

Query Match
Best Local Similarity 100.0%; Score 7; DB 6; Length 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 117 PLSTLEL 123
Db 5 PLSTLEL 11

RESULT 23
5171673-7
; Patent No. 5171673
; APPLICANT: SLOMA, ALAN; HANNETT, NANCY M.; STEPHENS, M.A.
; RUDOLPH, CATHY F.; RUFO JR., GERALD A.; PERO, JANICE
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS DNA USING
; THE BACILLUS COAGULANS AMYLASE GENE
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/219,599
; FILING DATE: 18-JUL-1988
; SEQ ID NO: 7:
; LENGTH: 35
5171673-7

Query Match
Best Local Similarity 100.0%; Score 7; DB 6; Length 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 166 KRLYARL 172
Db 5 KRLYARL 11

RESULT 24
US-09-149-476-601
; Sequence 601, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
```

1	EARLIER APPLICATION NUMBER: 60/040, 334
2	EARLIER FILING DATE: 1997-03-07
3	EARLIER APPLICATION NUMBER: 60/040, 336
4	EARLIER FILING DATE: 1997-03-07
5	EARLIER APPLICATION NUMBER: 60/040, 163
6	EARLIER FILING DATE: 1997-03-07
7	EARLIER APPLICATION NUMBER: 60/047, 600
8	EARLIER FILING DATE: 1997-05-23
9	EARLIER APPLICATION NUMBER: 60/047, 615
10	EARLIER FILING DATE: 1997-05-23
11	EARLIER APPLICATION NUMBER: 60/047, 597
12	EARLIER FILING DATE: 1997-05-23
13	EARLIER APPLICATION NUMBER: 60/047, 502
14	EARLIER FILING DATE: 1997-05-23
15	EARLIER APPLICATION NUMBER: 60/047, 633
16	EARLIER FILING DATE: 1997-05-23
17	EARLIER APPLICATION NUMBER: 60/047, 583
18	EARLIER FILING DATE: 1997-05-23
19	EARLIER APPLICATION NUMBER: 60/047, 617
20	EARLIER FILING DATE: 1997-05-23
21	EARLIER APPLICATION NUMBER: 60/047, 618
22	EARLIER FILING DATE: 1997-05-23
23	EARLIER APPLICATION NUMBER: 60/047, 503
24	EARLIER FILING DATE: 1997-05-23
25	EARLIER APPLICATION NUMBER: 60/047, 592
26	EARLIER FILING DATE: 1997-05-23
27	EARLIER APPLICATION NUMBER: 60/047, 581
28	EARLIER FILING DATE: 1997-05-23
29	EARLIER APPLICATION NUMBER: 60/047, 584
30	EARLIER FILING DATE: 1997-05-23
31	EARLIER APPLICATION NUMBER: 60/047, 500
32	EARLIER FILING DATE: 1997-05-23
33	EARLIER APPLICATION NUMBER: 60/047, 587
34	EARLIER FILING DATE: 1997-05-23
35	EARLIER APPLICATION NUMBER: 60/047, 492
36	EARLIER FILING DATE: 1997-05-23
37	EARLIER APPLICATION NUMBER: 60/047, 598
38	EARLIER FILING DATE: 1997-05-23
39	EARLIER APPLICATION NUMBER: 60/047, 613
40	EARLIER FILING DATE: 1997-05-23
41	EARLIER APPLICATION NUMBER: 60/047, 582
42	EARLIER FILING DATE: 1997-05-23
43	EARLIER APPLICATION NUMBER: 60/047, 596
44	EARLIER FILING DATE: 1997-05-23
45	EARLIER APPLICATION NUMBER: 60/047, 612
46	EARLIER FILING DATE: 1997-05-23
47	EARLIER APPLICATION NUMBER: 60/047, 632
48	EARLIER FILING DATE: 1997-05-23
49	EARLIER APPLICATION NUMBER: 60/047, 601
50	EARLIER FILING DATE: 1997-05-23
51	EARLIER APPLICATION NUMBER: 60/043, 580
52	EARLIER FILING DATE: 1997-04-11
53	EARLIER APPLICATION NUMBER: 60/043, 568
54	EARLIER FILING DATE: 1997-04-11
55	EARLIER APPLICATION NUMBER: 60/043, 314
56	EARLIER FILING DATE: 1997-04-11
57	EARLIER APPLICATION NUMBER: 60/043, 569
58	EARLIER FILING DATE: 1997-04-11
59	EARLIER APPLICATION NUMBER: 60/043, 311
60	EARLIER FILING DATE: 1997-04-11
61	EARLIER APPLICATION NUMBER: 60/043, 671
62	EARLIER FILING DATE: 1997-04-11
63	EARLIER APPLICATION NUMBER: 60/043, 674
64	EARLIER FILING DATE: 1997-04-11
65	EARLIER APPLICATION NUMBER: 60/043, 669
66	EARLIER FILING DATE: 1997-04-11
67	EARLIER APPLICATION NUMBER: 60/043, 312
68	EARLIER FILING DATE: 1997-04-11
69	EARLIER APPLICATION NUMBER: 60/043, 313
70	EARLIER FILING DATE: 1997-04-11
71	EARLIER APPLICATION NUMBER: 60/043, 672
72	EARLIER FILING DATE: 1997-04-11
73	EARLIER APPLICATION NUMBER: 60/043, 315
74	EARLIER FILING DATE: 1997-04-11
75	EARLIER APPLICATION NUMBER: 60/048, 974
76	EARLIER FILING DATE: 1997-06-06
77	EARLIER APPLICATION NUMBER: 60/056, 886
78	EARLIER FILING DATE: 1997-08-22
79	EARLIER APPLICATION NUMBER: 60/056, 877
80	EARLIER FILING DATE: 1997-08-22
81	EARLIER APPLICATION NUMBER: 60/056, 889
82	EARLIER FILING DATE: 1997-08-22
83	EARLIER APPLICATION NUMBER: 60/056, 893
84	EARLIER FILING DATE: 1997-08-22
85	EARLIER APPLICATION NUMBER: 60/056, 630
86	EARLIER FILING DATE: 1997-08-22
87	EARLIER APPLICATION NUMBER: 60/056, 878
88	EARLIER FILING DATE: 1997-08-22
89	EARLIER APPLICATION NUMBER: 60/056, 662
90	EARLIER FILING DATE: 1997-08-22
91	EARLIER APPLICATION NUMBER: 60/056, 872
92	EARLIER FILING DATE: 1997-08-22
93	EARLIER APPLICATION NUMBER: 60/056, 882
94	EARLIER FILING DATE: 1997-08-22
95	EARLIER APPLICATION NUMBER: 60/056, 637
96	EARLIER FILING DATE: 1997-08-22
97	EARLIER APPLICATION NUMBER: 60/056, 903
98	EARLIER FILING DATE: 1997-08-22
99	EARLIER APPLICATION NUMBER: 60/056, 888
100	EARLIER FILING DATE: 1997-08-22
101	EARLIER APPLICATION NUMBER: 60/056, 879
102	EARLIER FILING DATE: 1997-08-22
103	EARLIER APPLICATION NUMBER: 60/056, 880
104	EARLIER FILING DATE: 1997-08-22
105	EARLIER APPLICATION NUMBER: 60/056, 894
106	EARLIER FILING DATE: 1997-08-22
107	EARLIER APPLICATION NUMBER: 60/056, 911
108	EARLIER FILING DATE: 1997-08-22
109	EARLIER APPLICATION NUMBER: 60/056, 636
110	EARLIER FILING DATE: 1997-08-22
111	EARLIER APPLICATION NUMBER: 60/056, 874
112	EARLIER FILING DATE: 1997-08-22
113	EARLIER APPLICATION NUMBER: 60/056, 910
114	EARLIER FILING DATE: 1997-08-22
115	EARLIER APPLICATION NUMBER: 60/056, 864
116	EARLIER FILING DATE: 1997-08-22
117	EARLIER APPLICATION NUMBER: 60/

EARLIER APPLICATION NUMBER: 60/043,576  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/047,501  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,670  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/056,632  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,664  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,876  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 0.8%; Score 7; DB 4; Length 69;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 ELLEOLL 394  
DB 51 ELLEOLL 57

US-09-149-476-465  
Sequence 465, Application US/09149476  
Patent No. 6420526  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 186 Human Secreted proteins  
FILE REFERENCE: P2002pi  
CURRENT APPLICATION NUMBER: US/09/149,476  
CURRENT FILING DATE: 1998-09-08  
EARLIER APPLICATION NUMBER: PCT/US98/04493  
EARLIER FILING DATE: 1998-03-06  
EARLIER APPLICATION NUMBER: 60/040,162  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,333  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/038,621  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,626  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,334  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,336  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,163  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/047,600

EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,615  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,597  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,502  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,633  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,583  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,617  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,618  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,503  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,592  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,581  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,584  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,500  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,587  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,492  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,598  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,613  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,582  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,596  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,612  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,632  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,601  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,580  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,568  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,314  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,569  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,311  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,671  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,313  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,672  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,315  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/048,974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/056,886  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,877  
EARLIER FILING DATE: 1997-08-22

EARLIER APPLICATION NUMBER: 60/056,889  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,893  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,630  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,878  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,662  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,872  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,882  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,637  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,903  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,888  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,879  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,880  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,894  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,911  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,636  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,874  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,910  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,864  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,892  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,761  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/047,595  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,599  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,588  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,585  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,586  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,590  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,594  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,589  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,593  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,614  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,578  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,576  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/047,501  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,670  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/056,632

EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,664  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,876  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 0.8%; Score 7; DB 4; Length 70;  
Best local similarity 100.0%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 388 ELLEQLL 394  
Db 51 ELLEQLL 57

RESULT 26  
US-08-438-753B-24  
Sequence 24, Application US/08438753B  
Patent No. 5705363  
GENERAL INFORMATION:  
APPLICANT: Imakawa, Kazuhito  
TITLE OF INVENTION: Interferon Tau Compositions and  
METHODS OF USE  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESS: Dehlinger & Associates  
STREET: 350 Cambridge Ave., Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/438,753B  
FILING DATE: 10-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/139,891  
FILING DATE: 19-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/847,741  
FILING DATE: 09-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/318,050  
FILING DATE: 02-MAR-1989

```

983 6 0.7 517 1 MBEA_ECOLI p13658 escherichia
984 6 0.7 517 1 Y652_HUMAN 075143 homo sapien
985 6 0.7 518 1 ASB3_HUMAN 095175 homo sapien
986 6 0.7 519 1 CP5J_CANNA 012588 candida mai
987 6 0.7 519 1 DHA5_YEAST P40047 saccharomyc
988 6 0.7 520 1 CFP3_HUMAN Q08477 homo sapien
989 6 0.7 520 1 C2CB_ALCEU P13510 alcaigenes
990 6 0.7 520 1 C2CB_ALCEU P13510 alcaigenes
991 6 0.7 521 1 IMAL_XENLA P5170 xenopus lae
992 6 0.7 522 1 RECN_HELPT 092180 helicobacte
993 6 0.7 523 1 CYP4_CAEL P52012 caenorhabdi
994 6 0.7 523 1 U121_PPNV3 Q00703 pseudococc
995 6 0.7 524 1 G6PD_SYNP7 P29686 synecococc
996 6 0.7 524 1 NIKA_ECOLI P33590 escherichia
997 6 0.7 524 1 P60_LISTV 001837 listeria iv
998 6 0.7 524 1 RECN_HELPT 025943 helicobacte
999 6 0.7 525 1 ASB3_MOUSE 095172 mus musculu
1000 6 0.7 526 1 BUTY_HUMAN Q13410 homo sapien

```

## ALIGNMENTS

```

RESULT 1
YTFM_HAEIN STANDARD: PRT: 578 AA.
ID YTFM_HAEIN
AC P44038;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein HI0698 precursor.
GN HI0698.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
NCBI_TaxID=727;
[1]
SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RA MEDLINE=95305630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Scott J.D., Shirley R., Liu L.-I., Gilek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ulteback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RN Science 269:496-512(1995).
[2]
IDENTIFICATION BY MASS SPECTROMETRY.
RA MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RN Electrophoresis 21:411-429(2000).
CC -1- SIMILARITY: STRONG, TO E.COLI YTFM.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U32752; AAC2357.1;
DR TIGR: HI0698;
DR InterPro: IPR000184; Bac_surfAg-D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.

```

```

KW Signal: Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 578 PROTEIN HI0698.
SQ SEQUENCE 578 AA: 65812 MW: 34F9AC189C505876 CRC64;
Query Match 1.2%: Score 11; DB 1; Length 578;
Best Local Similarity 100.0%; Pred. No. 0.0072;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 882 GVGVRNAPVG 892
Db 539 GVGVRNAPVG 549

```

## RESULT 2

```

ID STA2_HUMAN STANDARD: PRT: 851 AA.
AC P52630; Q16430; Q16431;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Signal transducer and activator of transcription 2 (p113).
GN STAT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RA MEDLINE=95192056; PubMed=7885841;
RA Yan R., Qureshi S., Zhong Z., Wen Z., Darnell J.E. Jr.;
RT "The genomic structure of the STAT genes: multiple exons in
RT coincident sites in Stat1 and Stat2.";
RL Nucleic Acids Res. 23:459-463(1995).
[2]
SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RA MEDLINE=92366558; PubMed=1502204;
RA Fu X.-Y., Schindler C., Imbrota T., Aebersold R., Darnell J.E. Jr.;
RT "The proteins of ISGF-3, the interferon alpha-induced transcriptional
RT activator, define a gene family involved in signal transduction.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7840-7843(1992).
[3]
ALTERNATIVE SPLICING.
RA MEDLINE=96176320; PubMed=8601453;
RA Sugiyama T., Nishio Y., Kishimoto T., Akira S.;
RT "Identification of alternative splicing form of Stat2.";
RL FEBS Lett. 381:191-194(1996).
[4]
CHARACTERIZATION.
RP MEDLINE=97172544; PubMed=9020188;
RA Bluyssen H.A., Levy D.E.;
RT "Stat2 is a transcriptional activator that requires sequence-specific
RT contacts provided by stat1 and p48 for stable interaction with DNA.";
RN J Biol. Chem. 272:4600-4605(1997).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO THE IFN-STIMULATED
CC RESPONSE ELEMENT (ISRE) AND TO THE GAS ELEMENT. THIS MULTIPROTEIN
CC TRANSCRIPTION FACTOR IS TERMED ISG3.
CC -1- SUBUNIT: IN RESPONSE TO IFN ALPHA/BETA, THREE SUBUNITS (STAT1-
CC ALPHA, STAT1-BETA, STAT2) OF ISG3, BECOME PHOSPHORYLATED ON
CC TYROSINE, MIGRATE INTO THE NUCLEUS, AND ASSEMBLE INTO A COMPLEX
CC TOGETHER WITH ISG3 GAMMA (P48), A DNA-BINDING PROTEIN THAT
CC SPECIFICALLY BINDS TO THE IFN-STIMULATED RESPONSE ELEMENT.
CC -1- SUBCELLULAR LOCATION: Nuclear; translocated into the nucleus in
CC response to phosphorylation.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PTM: TYROSINE PHOSPHORYLATED IN RESPONSE TO IFN-ALPHA.
CC -1- SIMILARITY: BELONGS TO THE STAT FAMILY OF TRANSCRIPTION FACTORS.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```









399	6	0.7	281	1	ARAQ_SALTY	P03022 salmonella	472	6	0.7	313	1	YK89_ARCFU	O28190 archaeoglob
400	6	0.7	281	1	PANC_XYLFPA	O969r8 xyella fas	473	6	0.7	314	1	ETFA_BRAVA	P53573 bradyrhizob
401	6	0.7	281	1	Y125_MYCPN	P75511 mycoplasma	474	6	0.7	314	1	KHSE_HAEIN	P44504 haemophilus
402	6	0.7	282	1	RIP2_BRVDT	P98184 dryonia dlo	475	6	0.7	314	1	KHSE_PASMU	O94034 pasteurella
403	6	0.7	282	1	Y134_MERTA	O57598 methanococ	476	6	0.7	315	1	SPS1_BACSU	P39630 pasteurella
404	6	0.7	283	1	GAG1_HUMAN	P10263 homo sapien	477	6	0.7	316	1	COXT_YEAST	P53339 saccharomyc
405	6	0.7	283	1	HOXQ_AZOVI	P40594 azotobacter	478	6	0.7	316	1	GSEP_BACLI	P80057 bacillus li
406	6	0.7	283	1	TONB_NEIGO	O06432 neisseria g	479	6	0.7	317	1	HEW2_DEIRA	O94998 deinococcus
407	6	0.7	284	1	HUPQ_AZOC	O43957 azotobacter	480	6	0.7	317	1	HEW3_CIALE	P34797 clarkia lew
408	6	0.7	285	1	FLA3_TREMA	O949w9 treponema m	481	6	0.7	317	1	YVCK_BACSU	O06974 bacillus su
409	6	0.7	286	1	DAPE_BACHD	O947f0 bacillus ha	482	6	0.7	318	1	KHSE_VIBCH	O949k4 vibrio chol
410	6	0.7	286	1	PAL_KLEPN	P37446 klebsiella	483	6	0.7	319	1	YVCK_MYCLE	O949b3 bacillus ha
411	6	0.7	286	1	PAL_KLEPN	O94e87 rickettsia	484	6	0.7	319	1	TAL_RALSO	O84014 ralstonia s
412	6	0.7	286	1	RER2_YEAST	P35196 saccharomyc	485	6	0.7	319	1	YVCK_MYCLE	O84014 ralstonia s
413	6	0.7	287	1	Y791_CHLMU	O94jnt chlamydia m	486	6	0.7	320	1	HLPI_HUMAN	O949b3 bacillus ha
414	6	0.7	287	1	YICC_ECOLI	P23839 escherichia	487	6	0.7	321	1	LPD_CAME	O50098 mycobacteri
415	6	0.7	287	1	YICC_ECOLI	P44726 haemophilus	488	6	0.7	321	1	NECD_HUMAN	O84014 ralstonia s
416	6	0.7	288	1	YICC_HAEIN	P47182 saccharomyc	489	6	0.7	321	1	YVCK_MYCLE	O949b3 bacillus ha
417	6	0.7	288	1	Y504_CHLTR	O84512 chlamydia t	490	6	0.7	321	1	YVCK_MYCLE	O949b3 bacillus ha
418	6	0.7	289	1	THTR_CHICK	P25324 gallus gall	491	6	0.7	321	1	YVCK_MYCLE	O949b3 bacillus ha
419	6	0.7	291	1	RGR_BOVIN	P47803 bos taurus	492	6	0.7	322	1	YVCK_MYCLE	O949b3 bacillus ha
420	6	0.7	291	1	RGR_BOVIN	O94203 mus musculu	493	6	0.7	322	1	YVCK_MYCLE	O949b3 bacillus ha
421	6	0.7	292	1	ARAQ_ECOLI	P03021 escherichia	494	6	0.7	323	1	YVCK_MYCLE	O949b3 bacillus ha
422	6	0.7	292	1	E434_ADE09	P89083 human adeno	495	6	0.7	323	1	YVCK_MYCLE	O949b3 bacillus ha
423	6	0.7	292	1	KHSE_PYRPU	O84129 pyrococcus	496	6	0.7	323	1	YVCK_MYCLE	O949b3 bacillus ha
424	6	0.7	293	1	RL5_CAEEL	P49405 caenorhabdi	497	6	0.7	323	1	YVCK_MYCLE	O949b3 bacillus ha
425	6	0.7	293	1	ROCL_HUMAN	O60812 homo sapien	498	6	0.7	325	1	YVCK_MYCLE	O949b3 bacillus ha
426	6	0.7	294	1	E434_ADE02	P03239 human adeno	499	6	0.7	326	1	YVCK_MYCLE	O949b3 bacillus ha
427	6	0.7	294	1	NPM_CHICK	P16039 gallus gall	500	6	0.7	326	1	YVCK_MYCLE	O949b3 bacillus ha
428	6	0.7	295	1	RPSQ_MYXXA	O07083 myxococcus	501	6	0.7	326	1	YVCK_MYCLE	O949b3 bacillus ha
429	6	0.7	295	1	THTR_RAT	P24329 rattus norv	502	6	0.7	326	1	YVCK_MYCLE	O949b3 bacillus ha
430	6	0.7	296	1	DAPE_MYCLE	P46814 mycobacteri	503	6	0.7	327	1	YVCK_MYCLE	O949b3 bacillus ha
431	6	0.7	296	1	NPL_ECOLI	P06995 escherichia	504	6	0.7	327	1	YVCK_MYCLE	O949b3 bacillus ha
432	6	0.7	296	1	THTR_BOVIN	P00586 bos taurus	505	6	0.7	328	1	YVCK_MYCLE	O949b3 bacillus ha
433	6	0.7	296	1	THTR_CRIGR	P46635 escherichia	506	6	0.7	328	1	YVCK_MYCLE	O949b3 bacillus ha
434	6	0.7	296	1	YVCK_MYXXA	O55717 synechocyst	507	6	0.7	328	1	YVCK_MYCLE	O949b3 bacillus ha
435	6	0.7	296	1	YVCK_MYXXA	P578850 dictyostell	508	6	0.7	328	1	YVCK_MYCLE	O949b3 bacillus ha
436	6	0.7	297	1	YK46_METAC	O94449 thermotoga	509	6	0.7	329	1	YVCK_MYCLE	O949b3 bacillus ha
437	6	0.7	299	1	G3P_DICDI	O60382 methanococ	510	6	0.7	330	1	YVCK_MYCLE	O949b3 bacillus ha
438	6	0.7	299	1	MRAM_THEMA	O88792 mus musculu	511	6	0.7	330	1	YVCK_MYCLE	O949b3 bacillus ha
439	6	0.7	300	1	ARGW_MERJA	O24266 thermotoga	512	6	0.7	330	1	YVCK_MYCLE	O949b3 bacillus ha
440	6	0.7	300	1	JAM1_MOUSE	O60382 methanococ	513	6	0.7	330	1	YVCK_MYCLE	O949b3 bacillus ha
441	6	0.7	301	1	Y1J6_ECOLI	O88792 mus musculu	514	6	0.7	330	1	YVCK_MYCLE	O949b3 bacillus ha
442	6	0.7	301	1	Y1J6_ECOLI	P32667 caenorhabdi	515	6	0.7	331	1	YVCK_MYCLE	O949b3 bacillus ha
443	6	0.7	302	1	PTB_CLOBE	O05624 clostridium	516	6	0.7	331	1	YVCK_MYCLE	O949b3 bacillus ha
444	6	0.7	303	1	UI24_HSVSA	O01005 herpessvirus	517	6	0.7	332	1	YVCK_MYCLE	O949b3 bacillus ha
445	6	0.7	304	1	YB81_YEAST	P38326 saccharomyc	518	6	0.7	333	1	YVCK_MYCLE	O949b3 bacillus ha
446	6	0.7	304	1	MDH_HAUMA	O07841 halocaula	519	6	0.7	333	1	YVCK_MYCLE	O949b3 bacillus ha
447	6	0.7	305	1	NAC_ECOLI	O47005 escherichia	520	6	0.7	333	1	YVCK_MYCLE	O949b3 bacillus ha
448	6	0.7	305	1	PPP6_RAT	O64620 rattus norv	521	6	0.7	333	1	YVCK_MYCLE	O949b3 bacillus ha
449	6	0.7	305	1	Y348_MYCPN	P75255 mycoplasma	522	6	0.7	333	1	YVCK_MYCLE	O949b3 bacillus ha
450	6	0.7	306	1	PYRR_NEIMA	O949q5 neisseria m	523	6	0.7	333	1	YVCK_MYCLE	O949b3 bacillus ha
451	6	0.7	306	1	ROC_HUMAN	P07910 homo sapien	524	6	0.7	333	1	YVCK_MYCLE	O949b3 bacillus ha
452	6	0.7	306	1	YVCK_MYXXA	O67310 aquifex aeo	525	6	0.7	334	1	YVCK_MYCLE	O949b3 bacillus ha
453	6	0.7	308	1	FMT_CAUCR	O949eb9 caulobacter	526	6	0.7	335	1	YVCK_MYCLE	O949b3 bacillus ha
454	6	0.7	308	1	HCGP_MOUSE	O02614 mus musculu	527	6	0.7	337	1	YVCK_MYCLE	O949b3 bacillus ha
455	6	0.7	309	1	CD86_MOUSE	P42082 mus musculu	528	6	0.7	337	1	YVCK_MYCLE	O949b3 bacillus ha
456	6	0.7	309	1	KHSE_SALTY	O849p5 salmonella	529	6	0.7	338	1	YVCK_MYCLE	O949b3 bacillus ha
457	6	0.7	309	1	KHSE_SALTY	P27722 serralia ma	530	6	0.7	338	1	YVCK_MYCLE	O949b3 bacillus ha
458	6	0.7	309	1	KHSE_SALTY	O949b2 streptomyces	531	6	0.7	338	1	YVCK_MYCLE	O949b3 bacillus ha
459	6	0.7	309	1	KHSE_STRCO	O849n5 yersinia pe	532	6	0.7	338	1	YVCK_MYCLE	O949b3 bacillus ha
460	6	0.7	309	1	KHSE_YERPE	P11082 homo sapien	533	6	0.7	339	1	YVCK_MYCLE	O949b3 bacillus ha
461	6	0.7	309	1	MRAM_MYCCE	P29175 fibr murine	534	6	0.7	339	1	YVCK_MYCLE	O949b3 bacillus ha
462	6	0.7	310	1	P2AB_HUMAN	P11082 homo sapien	535	6	0.7	339	1	YVCK_MYCLE	O949b3 bacillus ha
463	6	0.7	310	1	GAC_MSYFR	O84982 escherichia	536	6	0.7	339	1	YVCK_MYCLE	O949b3 bacillus ha
464	6	0.7	310	1	KHSE_ECO57	O84982 escherichia	537	6	0.7	339	1	YVCK_MYCLE	O949b3 bacillus ha
465	6	0.7	310	1	KHSE_ECO57	O84982 escherichia	538	6	0.7	339	1	YVCK_MYCLE	O949b3 bacillus ha
466	6	0.7	311	1	MRAM_CLOPE	O84982 escherichia	539	6	0.7	340	1	YVCK_MYCLE	O949b3 bacillus ha
467	6	0.7	311	1	PYDB_LACLA	O84982 escherichia	540	6	0.7	341	1	YVCK_MYCLE	O949b3 bacillus ha
468	6	0.7	311	1	PYDB_LACLA	O84982 escherichia	541	6	0.7	341	1	YVCK_MYCLE	O949b3 bacillus ha
469	6	0.7	312	1	YVCK_MYCLE	O84982 escherichia	542	6	0.7	341	1	YVCK_MYCLE	O949b3 bacillus ha
470	6	0.7	313	1	YVCK_MYCLE	O84982 escherichia	543	6	0.7	341	1	YVCK_MYCLE	O949b3 bacillus ha
471	6	0.7	313	1	YVCK_MYCLE	O84982 escherichia	544	6	0.7	341	1	YVCK_MYCLE	O949b3 bacillus ha





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: April 28, 2003, 16:24:59 ; Search time 16 Seconds  
(Without alignments)  
2382.297 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 919  
Sequence: 1 MSKPVLFANRSEMPVALAAV.....TGVKEEGNPILKHFITGTPP 919

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt-40:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	1.2	578	1 YTFM_HAETN	P44038 haemophilus
2	9	1.0	851	1 STRA2_HUMAN	P52630 homo sapien
3	8	0.9	200	1 NAPC_ECOLI	P31932 escherichia
4	8	0.9	360	1 KAG2_CAEEL	Q27535 caenorhabdi
5	8	0.9	577	1 YTFM_ECOLI	P39320 escherichia
6	8	0.9	734	1 MCCA_ARATH	O45223 arabidopsis
7	8	0.9	1047	1 CARB_THENC	O9HK17 thermoplas
8	8	0.9	1355	1 SALM_DROME	P38770 drosophila
9	8	0.9	1402	1 SALM_DROME	P38066 drosophila
10	7	0.8	82	1 HYBG_ECOLI	P37185 escherichia
11	7	0.8	107	1 YGT2_YEAST	P53103 saccharomyc
12	7	0.8	109	1 Y353_MYCPN	P75249 mycoplasma
13	7	0.8	159	1 MP63_MYCTU	P97175 mycobacteri
14	7	0.8	190	1 GRA1_TOXCO	P13403 toxoplasma
15	7	0.8	190	1 RS7_AVIMR	O9ZNS1 avicennia m
16	7	0.8	195	1 INO1_BOVIN	P07352 bos taurus
17	7	0.8	211	1 PDOI_ECOLI	P32174 escherichia
18	7	0.8	211	1 FIXJ_AZOCA	P26487 azorhizobiu
19	7	0.8	223	1 VATD_THERM	O87880 thermus the
20	7	0.8	226	1 GPH_VIBCH	O9KNV6 vibrio chol
21	7	0.8	232	1 COX2_ASCSU	P24882 ascaris suu
22	7	0.8	242	1 RSTA_HUMAN	P52108 escherichia
23	7	0.8	261	1 CLDI_HUMAN	P56857 homo sapien
24	7	0.8	264	1 CLDI_MOUSE	P56857 homo sapien
25	7	0.8	271	1 PURR_LACIA	O53065 lactococcus
26	7	0.8	280	1 LIGE_PSEPA	P27457 pseudomonas
27	7	0.8	280	1 YTTT_BACSU	P39803 bacillus su
28	7	0.8	296	1 THTR_HUMAN	Q16762 homo sapien
29	7	0.8	296	1 THTR_MOUSE	P52196 mus musculi
30	7	0.8	303	1 34KD_MYCTU	P71556 mycobacteri
31	7	0.8	313	1 OYE6_HUMAN	O966f5 homo sapien
32	7	0.8	314	1 MRAW_FUSNM	O86f65 fusobacteri
33	7	0.8	317	1 MOCA_RHIME	P49307 rhizobium m

34	7	0.8	323	1 ISPB_ECOLI	P19641 escherichia
35	7	0.8	331	1 EL3K_TOBAC	P52398 nicotiana t
36	7	0.8	332	1 HITTA_HAETN	P33755 haemophilus
37	7	0.8	347	1 YPFG_ECOLI	P76559 escherichia
38	7	0.8	349	1 GP21_HUMAN	O99679 homo sapien
39	7	0.8	358	1 HRCA_CAUCR	P54305 caulobacter
40	7	0.8	375	1 PCE_TACMR	P21902 tachypleus
41	7	0.8	383	1 Y628_MYCTU	P66918 mycobacteri
42	7	0.8	386	1 ALR_ANASP	O8Y96 anabaena sp
43	7	0.8	397	1 WH42_STRCO	P23158 streptomyc
44	7	0.8	398	1 GDN_HUMAN	P07093 homo sapien
45	7	0.8	402	1 YDCM_ECOLI	P76102 escherichia
46	7	0.8	404	1 PRHA_METH	P19496 methanobact
47	7	0.8	405	1 DGT1_AGR5	O8uep3 agrobacteri
48	7	0.8	423	1 TTG_BACSU	P80698 bacillus su
49	7	0.8	428	1 C124_MYCTU	O50696 mycobacteri
50	7	0.8	433	1 YBBY_ECOLI	P77328 escherichia
51	7	0.8	454	1 YB48_MYCTU	O06548 mycobacteri
52	7	0.8	454	1 YJ45_MYCTU	P95269 mycobacteri
53	7	0.8	481	1 INX6_DROME	O9YR82 drosophila
54	7	0.8	488	1 TFS1_KLULA	Q07158 kluyveromyc
55	7	0.8	496	1 SRM_MOUSE	P39976 saccharomyc
56	7	0.8	496	1 SRM_MOUSE	Q62270 mus musculi
57	7	0.8	498	1 FSC3_HUMAN	O9qut6 homo sapien
58	7	0.8	507	1 MOSB_RHIME	O07608 rhizobium m
59	7	0.8	512	1 AMY_BACLI	P06278 bacillus li
60	7	0.8	513	1 YMT3_YEAST	O04305 saccharomyc
61	7	0.8	519	1 AMYG_SACFI	P08017 saccharomyc
62	7	0.8	519	1 AMYG_SACFI	P26989 saccharomyc
63	7	0.8	522	1 MSAB_NEIMO	P14930 n peptide m
64	7	0.8	522	1 MSAB_NEIMA	O913m8 n peptide m
65	7	0.8	522	1 MSAB_NEIMA	O9K1n8 n peptide m
66	7	0.8	552	1 Y664_HAETN	O5738 haemophilus
67	7	0.8	567	1 IF37_SCHPO	O94236 schizosacch
68	7	0.8	571	1 YB63_SCHPO	Q09744 schizosacch
69	7	0.8	574	1 YHGE_ECOLI	P45804 escherichia
70	7	0.8	579	1 GPC2_RAT	P51653 rattus norv
71	7	0.8	587	1 DSD2_PSEAE	O91104 pseudomonas
72	7	0.8	595	1 REB1_KLULA	O05950 kluyveromyc
73	7	0.8	613	1 ILVD_YIBCH	O9K4W0 vibrio chol
74	7	0.8	624	1 AMYG_ARXAD	P42042 arxula aden
75	7	0.8	628	1 DNLJ_THEMA	O9KWX5 thermotoga
76	7	0.8	725	1 HS9B_BRARE	O57521 brachydanio
77	7	0.8	776	1 YOK5_CAEEL	O11177 caenorhabdi
78	7	0.8	807	1 MCM3_XENLA	P49739 xenopus lae
79	7	0.8	808	1 MCM3_HUMAN	P25206 homo sapien
80	7	0.8	812	1 YR86_CAEEL	O09568 caenorhabdi
81	7	0.8	830	1 HMT1_SCHPO	O02592 schizosacch
82	7	0.8	1026	1 EXSB_CHLMO	O9p1t8 chlamydia m
83	7	0.8	1068	1 P11A_BOVIN	P32871 bos taurus
84	7	0.8	1068	1 P11A_HUMAN	P42336 homo sapien
85	7	0.8	1068	1 P11A_MOUSE	P42337 mus musculi
86	7	0.8	1068	1 YML7_YEAST	O03725 saccharomyc
87	7	0.8	1134	1 SLAP_BACST	P35825 bacillus st
88	7	0.8	1228	1 NRCA_CHICK	P35331 gallus gall
89	7	0.8	1284	1 FRPA_NEIMB	O9K0K9 neisseria m
90	7	0.8	1302	1 XDH_DROME	P10351 drosophila
91	7	0.8	1335	1 XDH_DROPS	P22811 drosophila
92	7	0.8	1342	1 XDH_DROPS	P21711 drosophila
93	7	0.8	1444	1 XDH_DROPS	P39059 homo sapien
94	7	0.8	1388	1 CA1E_HUMAN	O57661 pyrococcus
95	7	0.8	1431	1 RGSC_HUMAN	O14924 homo sapien
96	7	0.8	1447	1 CHD2_HUMAN	O14647 homo sapien
97	7	0.8	1739	1 FRPC_NEIMB	O9JYV5 neisseria m
98	7	0.8	1829	1 FRPC_NEIMC	P15127 neisseria m
99	7	0.8	1829	1 GP21_RAT	P16544 rattus norv
100	7	0.8	2142	1 BAT2_HUMAN	P48634 homo sapien
101	7	0.8	2336	1 CCAB_RAT	O02284 rattus norv
102	7	0.8	3415	1 POLG_POWVL	O04538 t genome po
103	7	0.8	3436	1 CY35_DESAC	P81079 desulfuromo
104	6	0.7	46	1 LA89_LACAC	O48501 lactobacilli
105	6	0.7	48	1 ATP8_SCHPO	P21536 schizosacch
106	6	0.7	48	1 ATP8_SCHPO	



---



A48075  
60K stress-related protein - chicken (fragments)  
C:Species: Gallus gallus (chicken)  
C:Date: 16-Feb-1994 #sequence\_revision 07-Feb-1997 #text\_change 11-Jan-2000  
C:Accession: A48075; B48075; D48075; E48075; F48075; G48075  
R:Smith, D.F.; Sullivan, W.P.; Marlon, T.N.; Zaltsu, K.; Madden, B.; McCormick, D.J.; TC  
Mol. Cell. Biol. 13, 869-876, 1993  
A:Title: Identification of a 60-kilodalton stress-related protein, p60, which interacts  
A:Reference number: A48075; MUID:93140785; PMID:8423808  
A:Accession: A48075  
A:Molecule type: protein  
A:Residues: 1-36 <SM1>  
A:Note: sequence extracted from NCBI backbone (NCBIP:123364)  
A:Accession: B48075  
A:Molecule type: protein  
A:Residues: 37-43 <SM2>  
A:Note: sequence extracted from NCBI backbone (NCBIP:123366)  
A:Accession: C48075  
A:Molecule type: protein  
A:Residues: 44-59 <SM3>  
A:Note: sequence extracted from NCBI backbone (NCBIP:123368)  
A:Accession: D48075  
A:Molecule type: protein  
A:Residues: 60-73 <SM4>  
A:Note: sequence extracted from NCBI backbone (NCBIP:123370)  
A:Accession: E48075  
A:Molecule type: protein  
A:Residues: 74-98 <SM5>  
A:Note: sequence extracted from NCBI backbone (NCBIP:123372)  
A:Accession: F48075  
A:Molecule type: protein  
A:Residues: 99-108 <SM6>  
A:Note: sequence extracted from NCBI backbone (NCBIP:123374)  
A:Accession: G48075  
A:Molecule type: protein  
A:Residues: 109-124 <SM7>  
A:Note: sequence extracted from NCBI backbone (NCBIP:123376)  
C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat hc

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 124;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 387 RELLEQL 393  
|||||||  
Db 52 RELLEQL 58

RESULT 38  
F97593  
50S ribosomal protein L22 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: F97593  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194  
A:Accession: F97593  
A:Molecule type: DNA  
A:Residues: 1-129 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK87703.1; PID:915157064; GSPDB:GN00169  
A:Gene: AGR\_C\_3548  
A:Map position: circular chromosome  
C:Superfamily: Escherichia coli ribosomal protein L22

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 129;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 491 KLNLYAA 497  
|||||||  
Db 28 KLNLYAA 34

RESULT 39  
AC2815  
50S ribosomal protein L22 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002  
C:Accession: AC2815  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; MCCI  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AC2815  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-129 <KUR>  
A:Cross-references: GB:AE00688; PIDN:AA142937.1; PID:917740394; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: rplY  
A:Map position: circular chromosome  
C:Superfamily: Escherichia coli ribosomal protein L22

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 129;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 491 KLNLYAA 497  
|||||||  
Db 28 KLNLYAA 34

RESULT 40  
AC1545  
hypothetical protein lin0899 [imported] - Listeria innocua (strain Clp11262)  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AC1545  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitouram, A.;  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AC1545  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-130 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC96131.1; PID:916413349; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: lin0899

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 130;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 698 TOAPPET 704  
|||||||  
Db 15 TOAPPET 21

Search completed: April 28, 2003, 16:36:03  
Job time : 78 secs

submitted to the EMBL Data Library, May 1999  
A:Reference number: Z18806  
A:Accession: T18004  
A:Status: preliminary; translated from GB/EMBL/DDAJ

A:Molecule type: DNA  
A:Residues: 1-95 <GRA>  
A:Cross-references: EMBL:U42580; NID:q4028896; PIDN:AAC96869.1

A:Experimental source: specific host *Chlorella* strain NC64A  
C:Genetics:  
A:Note: A502L

Query Match 0.8%; Score 7; DB 2; Length 95;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 817 QSRGYA 823  
|||||  
Db 47 QSRGYA 53

RESULT 33  
S61135  
hypothetical protein YGL182c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein G1607  
C:Species: *Saccharomyces cerevisiae*

C>Date: 23-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 19-Apr-2002

C:Accession: S61135; S64199  
R:Bertani, I.; Cogillevina, M.; Zaccaria, P.; Klima, R.; Brusch, C.V.

R:Bertani, C.V.; Cogillevina, M.; Bertani, I.; Klima, R.; Zaccaria, P.; Delneri, D.  
submitted to the EMBL Data Library, September 1995

A:Description: The sequence analysis of a 7.9 kb DNA fragment from the left arm of *S.cerevisiae* new genes.

A:Reference number: S61138  
A:Accession: S61135

A:Molecule type: DNA  
A:Residues: 1-107 <BER>

A:Cross-references: EMBL:X91489; NID:q1143557; PIDN:CAA62792.1; PID:q1143565  
R:Brusch, C.V.; Cogillevina, M.; Bertani, I.; Klima, R.; Zaccaria, P.; Delneri, D.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64183  
A:Accession: S64199

A:Molecule type: DNA  
A:Residues: 1-107 <BRU>

A:Cross-references: EMBL:Z72704; NID:q1322794; PIDN:CAA96894.1; PID:q1322795; MIPS:YGL18  
A:Experimental source: strain S288C

C:Genetics:  
A:Cross-references: SGD:S0003150

A:Map position: 7L  
C:Superfamily: *Saccharomyces* hypothetical protein YGL182c

Query Match 0.8%; Score 7; DB 2; Length 107;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 687 YSLRYRL 693  
|||||  
Db 84 YSLRYRL 90

RESULT 34

S73639  
MG353 homolog g12-orf109 - *Mycoplasma pneumoniae* (strain ATCC 29342)

C:Species: *Mycoplasma pneumoniae*  
A:Variate: ATCC 29342

C>Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999

C:Accession: S73639  
R:Himmelfrich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.

Nucleic Acids Res. 24, 4420-4449, 1996

A:Title: Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*

A:Reference number: S73327; MUID:97105885; PMID:8948633

A:Accession: S73639  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-109 <HIM>

A:Cross-references: EMBL:AE000028; GB:U00089; NID:q1673972; PIDN:AA95961.1; PID:q167  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
C:Genetics:  
A:Genetic code: SGC3

Query Match 0.8%; Score 7; DB 2; Length 109;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 898 VATGYKE 904  
|||||  
Db 22 VATGYKE 28

RESULT 35  
A69441  
conserved hypothetical protein AF1530 - *Archaeoglobus fulgidus*

C:Species: *Archaeoglobus fulgidus*  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000

C:Accession: A69441  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A>Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AC00884  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-82 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD02975.1; PID:q16504223; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY3314  
 C:Superfamily: hydrogenase expression/formation protein hycp

Query Match 0.8%; Score 7; DB 2; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 839 GOVLAVG 845  
 Db 7 GOVLAVG 13

RESULT 28  
 C85958  
 hypothetical protein hycg [imported] - *Escherichia coli* (strain O157:H7, substrain EDL93  
 C:Species: *Escherichia coli*  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-May-2002  
 C:Accession: C85958  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 11ler, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A:Reference number: AB5480; MUID:21074935; PMID:11206551  
 A:Accession: C85958  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-82 <STO>  
 A:Cross-references: GB:AE005174; MID:q12517554; PIDN:AAG58127.1; GSPDB:GN00145; UWGP:Z43  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: hycg  
 C:Superfamily: hydrogenase expression/formation protein hycp

Query Match 0.8%; Score 7; DB 2; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 839 GOVLAVG 845  
 Db 7 GOVLAVG 13

RESULT 29  
 C91113  
 hydrogenase-2 operon protein hycg [imported] - *Escherichia coli* (strain O157:H7, substr  
 C:Species: *Escherichia coli*  
 C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 17-May-2002  
 C:Accession: C91113  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 gawara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genc  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: C91113  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-82 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA37298.1; PID:q13363347; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RMD 0509952  
 C:Genetics:  
 A:Gene: ECS3875  
 C:Superfamily: hydrogenase expression/formation protein hycp

Query Match 0.8%; Score 7; DB 2; Length 82;

Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 839 GOVLAVG 845  
 Db 7 GOVLAVG 13

RESULT 30  
 AH0088  
 probable flagellar regulatory protein YPO0720 [imported] - *Yersinia pestis* (strain CO  
 C:Species: *Yersinia pestis*  
 C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
 C:Accession: AH0088  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M  
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G  
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel  
 Nature 413, 523-527, 2001  
 A>Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:11586360  
 A:Accession: AH0088  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-88 <KUR>  
 A:Cross-references: GB:AL590842; PIDN:CAC89571.1; PID:q15978804; GSPDB:GN00175  
 C:Genetics:  
 A:Gene: YPO0720

Query Match 0.8%; Score 7; DB 2; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 533 AVARAIL 539  
 Db 77 AVARAIL 83

RESULT 31  
 AB1784  
 hypothetical protein NMA2123 [imported] - *Neisseria meningitidis* (strain Z2491 serogr  
 C:Species: *Neisseria meningitidis*  
 C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
 C:Accession: AB1784  
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
 ; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre  
 Nature 404, 502-506, 2000  
 A>Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491  
 A:Reference number: AB1775; MUID:20222556; PMID:10761919  
 A:Accession: AB1784  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-91 <PAR>  
 A:Cross-references: GB:AL162758; GB:AL157959; MID:q7380672; PIDN:CAB85336.1; PID:q738  
 A:Experimental source: serogroup A, strain Z2491  
 C:Genetics:  
 A:Gene: NMA2123

Query Match 0.8%; Score 7; DB 2; Length 91;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 LRPEQIO 67  
 Db 8 LRPEQIO 14

RESULT 32  
 T18004  
 hypothetical protein A502L - *Chlorella virus PBCV-1*  
 C:Species: *Chlorella virus PBCV-1*  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T18004  
 R:Graves, M.V.; Van Etten, J.L.

C:Accession: T00661  
 R:Federalipil, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo, I.; Vysotskaya, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.  
 submitted to the EMBL Data Library, February 1998  
 A:Reference number: Z14197  
 A:Accession: T00661  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1417 <PED>  
 A:Cross-references: EMBL:AC002396; NID:g2749918; PID:g2829883; GSPDB:GN00059; ATSP:F316.  
 C:Genetics:  
 A:Gene: ATSP:F316.24  
 A:Map position: 1  
 A:Introns: 23/1; 46/3; 276/1; 309/1; 546/1; 1269/1; 1284/2; 1333/1  
 Query Match  
 Best Local Similarity 100.0%; Score 8; DB 2; Length 1417;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 387 RELLEQL 394  
 |||||  
 Db 843 RELLEQL 850

RESULT 24  
 G82667  
 hypothetical protein XF1543 [imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: G82667  
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: G82667  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-66 <SIM>  
 A:Cross-references: GB:AE003984; GB:AE003849; NID:g9106581; PIDN:AAE4352.1; GSPDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tsunako, M.H.; Valada, H.; Van Sluys, W.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF1543

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 2; Length 66;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 809 LRFAGG 815  
 |||||  
 Db 11 LRFAGG 17

RESULT 25  
 S75519  
 hypothetical protein ssc7201 - Synechocystis sp. (strain PCC 6803)  
 C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C:Accession: S75519  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O.K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys  
 s.  
 A:Reference number: S74322; MUID:97061201; PMID:8905231  
 A:Accession: S75519  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-72 <KAN>  
 A:Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BA18080.1; PID:dl01  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 Query Match  
 Best Local Similarity 100.0%; Score 7; DB 2; Length 72;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 216 LEDITQE 222  
 |||||  
 Db 54 LEDITQE 60

RESULT 26  
 G55516  
 hydrogenase-2 operon protein hylB - Escherichia coli (strain K-12)  
 C:Species: Escherichia coli  
 C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 01-Mar-2002  
 C:Accession: G55516; D65085  
 R:Memon, N.K.; Chetels, C.Y.; Dervartanian, M.; Wendt, J.C.; Shammugam, K.T.; Peck J  
 J. Bacteriol. 176, 4416-4423, 1994  
 A:Title: Cloning, sequencing, and mutational analysis of the hylB operon encoding Esch  
 A:Reference number: A55516; MUID:94292472; PMID:8021226  
 A:Accession: G55516  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-82 <MEN>  
 A:Cross-references: GB:U09177; NID:g501051; PIDN:AAA21595.1; PID:g544489  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617; PMID:9278503  
 A:Accession: D65085  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-82 <BLAT>  
 A:Cross-references: GB:AE000362; GB:U00096; NID:g2367182; PIDN:AACT6026.1; PID:g17693  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:  
 A:Gene: hylB  
 C:Superfamily: hydrogenase expression/formation protein hylC

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 2; Length 82;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 839 GOVLAVG 845  
 |||||  
 Db 7 GOVLAVG 13

RESULT 27  
 AG0884  
 hydrogenase-2 component protein [imported] - Salmonella enterica subsp. enterica sero  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 17-May-2002  
 C:Accession: AG0884  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
 th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001

A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A:Reference number: AB5480; MUID:21074935; PMID:11206551  
A:Accession: F86119  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-577 <STO>  
A:Cross-references: GB:AE005174; NID:g12519219; PIDN:AG69418.1; GSPDB:GN00145; UMGp:258  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: ytfm

Query Match 0.9%; Score 8; DB 2; Length 577;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 809 LRFPGGD 816  
|||||||  
DB 466 LRFPGGD 473

RESULT 19  
F82064

conserved hypothetical protein VC2548 [imported] - *Vibrio cholerae* (strain N16961 serogr  
C:Species: *Vibrio cholerae*  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: F82064  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
charlson, D.; Ermolaeva, M.D.; Vamtheman, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F  
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A:Reference number: A82035; MUID:20406833; PMID:10952301  
A:Accession: F82064  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-582 <HEI>  
A:Cross-references: GB:AE004323; GB:AE003852; NID:g9657119; PIDN:AF95689.1; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC2548  
A:Map position: 1

Query Match 0.9%; Score 8; DB 2; Length 582;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 809 LRFPGGD 816  
|||||||  
DB 473 LRFPGGD 480

RESULT 20  
AH2546

hypothetical protein all7659 [imported] - *Nostoc* sp. (strain PCC 7120) plasmid PCC7120P6  
C:Species: *Nostoc* sp.  
A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120  
C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AH2546  
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iriuch  
Nakazaki, N.; Shimoto, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*  
A:Reference number: AB1807; MUID:21595285; PMID:11755840  
A:Accession: AH2546  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-731 <KUR>  
A:Cross-references: GB:AP003602; PIDN:BAW77302.1; PID:g17134744; GSPDB:GN00181  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all7659  
A:Genome: plasmid

Query Match 0.9%; Score 8; DB 2; Length 731;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 IKALEDI 219  
|||||||  
DB 116 IKALEDI 123

RESULT 21  
S40022

spalt protein - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C>Date: 19-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 21-Jul-2000  
C:Accession: S40022  
R:Kuehnlein, R.P.; Frommer, G.; Friedrich, M.; Gonzalez-Gaitan, M.; Weber, A.; Wagner  
EMBO J. 13, 168-179, 1994  
A:Title: spalt encodes an evolutionarily conserved zinc finger protein of novel struc  
A:Reference number: S40022; MUID:94139659; PMID:7905822  
A:Accession: S40022  
A:Molecule type: DNA  
A:Residues: 1-1355 <KUE>  
A:Cross-references: EMBL:X75541; NID:g414106; PIDN:CA53229.1; PID:g2598394  
A:Note: the authors translated the codon GAC for residue 51 as Ala  
A:Gene: sal; spalt  
A:Cross-references: FlyBase:FBgn0004579  
A:Introns: 51/1; 1329/3 1355/2  
C:Keywords: zinc finger

Query Match 0.9%; Score 8; DB 2; Length 1355;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LTFEIOA 68  
|||||||  
DB 913 LTFEIOA 920

RESULT 22  
S42748

finger protein - fruit fly (*Drosophila virilis*) (fragment)  
C:Species: *Drosophila virilis*  
C>Date: 20-Oct-1994 #sequence\_revision 26-May-1995 #text\_change 24-Sep-1998  
C:Accession: S42748  
R:Schuh, R.  
submitted to the EMBL Data Library, November 1993  
A:Reference number: S42748  
A:Accession: S42748  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1402 <SCH>  
A:Cross-references: EMBL:Z27444; NID:g426461; PID:g426462  
C:Genetics:  
A:Gene: FlyBase:FlyBase:FBgn0013137  
A:Cross-references: FlyBase:FBgn0013137  
C:Keywords: DNA binding; transcription regulation

Query Match 0.9%; Score 8; DB 2; Length 1402;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LTFEIOA 68  
|||||||  
DB 934 LTFEIOA 941

RESULT 23  
T00661

hypothetical protein F316.24 - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 22-Oct-1999

Best Local Similarity 100.0%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 492 LNLVAKA 499  
|||||||  
DB 254 LNLVAKA 261

# RESULT 14

JC5747  
coronafacic acid synthetase component cfa3 [imported] - Pseudomonas syringae

C:Species: Pseudomonas syringae  
C:Date: 24-Jan-1998 #sequence\_revision 13-Mar-1998 #text\_change 20-Jun-2000

C:Accession: JC5747  
R:Penfold, C.N.; Bender, C.L.; Turner, J.G.  
Gene 183, 167-173, 1996

A:Title: Characterisation of genes involved in biosynthesis of coronafacic acid, the pol

A:Reference number: JC5745; MUID:97149295; PMID:8996103

A:Accession: JC5747

A:Molecule type: DNA

A:Residues: 1-380 <PEN>  
A:Cross-references: GB:U56980; NID:91655810; PIDN:AAB41300.1; PID:91655813

A:Note: the authors translated the initiation codon GTG for residue 1 as Met

C:Genetics:  
A:Gene: cfa3  
A:Start codon: GTG  
C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I; 3-oxoacyl-[acyl-carrier-prot

C:Keywords: glycosidase; hydrolase; transferase  
F:26-372/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>

Query Match  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 529 DAVSAVAR 536  
|||||||  
DB 230 DAVSAVAR 237

# RESULT 15

T34645  
hypothetical protein SC10H5.07 SC10H5.07 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999

C:Accession: T34645  
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1998

A:Reference number: 221550

A:Accession: T34645

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-469 <OLI>  
A:Cross-references: EMBL:AL031232; PIDN:CAA20279.1; GSPDB:GN00070; SCODEB:SC10H5.07

A:Experimental source: strain A3(2)

C:Genetics:  
A:Gene: SCODEB:SC10H5.07

Query Match  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 PVLAAYL 21  
|||||||  
DB 420 PVLAAYL 427

# RESULT 16

S56445  
hypothetical 64.8k protein (msra-chpbl intergenic region) - Escherichia coli (strain K-1

N:Alternate names: hypothetical protein c577  
C:Species: Escherichia coli  
C:Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 01-Mar-2002

C:Accession: S56445; G65233  
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
Nucleic Acids Res. 23, 2105-2119, 1995

A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from

A:Reference number: S56314; MUID:95334362; PMID:7610040

A:Accession: S56445

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-577 <BUR>  
A:Cross-references: EMBL:U14003; NID:91263172; PIDN:AAA97116.1; PID:9537061

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G65233

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-577 <BLAT>  
A:Cross-references: GB:AE000493; GB:U00096; NID:92367360; PIDN:AACT7177.1; PID:917906

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:  
A:Gene: ytfM  
A:Start codon: GTG

Query Match  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 809 LRFFAGD 816  
|||||||  
DB 466 LRFFAGD 473

# RESULT 17

F91278  
hypothetical protein ECs5198 [imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: F91278

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9

A:Reference number: A9629; MUID:21156231; PMID:11258796

A:Accession: F91278

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-577 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA838621.1; PID:913364675; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:  
A:Gene: ECs5198

Query Match  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 809 LRFFAGD 816  
|||||||  
DB 466 LRFFAGD 473

# RESULT 18

F86119  
hypothetical protein ytfM [imported] - Escherichia coli (strain O157:H7, substrain ED

C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: F86119

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dialantana, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001

RESULT 9  
E85859  
cytochrome c-type protein [imported] - Escherichia coli (strain O157:H7, substrain EDL93)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: E85859  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: AB5480; MUID:21074935; PMID:11206551  
A:Accession: E85859  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-200 <STO>  
A:Cross-references: GB:AE005174; NID:q12516534; PIDN:AAG57337.1; GSPDB:GN00145; UWGP:234  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: napC  
C:Superfamily: denitrification system component nirt; nirt homology

Query Match 0.9%; Score 8; DB 2; Length 200;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 161 KPGLIKRL 168  
Db 7 KPGLIKRL 14

RESULT 10  
AE1390  
MreB-like protein homolog mbl [imported] - Listeria monocytogenes (strain EGD-e)  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AE1390  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,  
D.; Jones, L.M.; Karst, U.  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AE1390  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-331 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAD00603.1; PID:q16412013; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: mbl  
C:Superfamily: rod shape-determining protein envB

Query Match 0.9%; Score 8; DB 2; Length 331;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 839 GOVLAVGT 846  
Db 38 GOVLAVGT 45

RESULT 11  
AG1765  
MreB-like protein homolog mbl [imported] - Listeria innocua (strain Cl1p11262)  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AG1765  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.  
D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maltournam, A.;  
Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A:Title: Comparative genomics of Listeria species  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AG1765  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-331 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC97895.1; PID:q16415205; GSPDB:GN00178  
A:Experimental source: strain Cl1p11262  
C:Genetics:  
A:Gene: mbl  
C:Superfamily: rod shape-determining protein envB

Query Match 0.9%; Score 8; DB 2; Length 331;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 839 GOVLAVGT 846  
Db 38 GOVLAVGT 45

RESULT 12  
T44739  
hypothetical protein MLCB1450.24 [imported] - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000  
C:Accession: T44739  
R:James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
Submitted to the EMBL Data Library, January 1998  
A:Reference number: Z22831  
A:Accession: T44739  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-340 <JAM>  
A:Cross-references: EMBL:AL035159; PIDN:CAA22708.1  
A:Experimental source: cosmid B1450  
C:Genetics:  
A:Note: MLCB1450.24

Query Match 0.9%; Score 8; DB 2; Length 340;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 884 GVRWASPV 891  
Db 167 GVRWASPV 174

RESULT 13  
T27569  
hypothetical protein ZC434.8 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T27569  
R:Milkinson, J.  
Submitted to the EMBL Data Library, July 1996  
A:Reference number: Z20388  
A:Accession: T27569  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-360 <WIL>  
A:Cross-references: EMBL:Z75714; PIDN:CAB00062.1; GSPDB:GN00019; CESP:ZC434.8  
A:Experimental source: clone ZC434  
C:Genetics:  
A:Gene: CESP:ZC434.8  
A:Map position: 1  
A:Introns: 13/3; 109/3; 310/3  
C:Superfamily: creatine kinase; creatine kinase repeat homology

Query Match 0.9%; Score 8; DB 2; Length 360;

A46160  
interferon alpha-induced transcription activator ISGF-3, 113K chain - human  
N:Alternate names: stat2 protein  
C:Species: Homo sapiens (man)  
C>Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Aug-2000  
C:Accession: A46160; S71908; S53873  
R:Fu, X.Y.; Schindler, C.; Improt, T.; Abersold, R.; Darnell Jr., J.E.  
Proc. Natl. Acad. Sci. U.S.A. 89, 7840-7843, 1992  
A:Title: The proteins of ISGF-3, the interferon alpha-induced transcriptional activator,  
A:Reference number: A46160; MUID:92366558; PMID:1502204  
A:Accession: A46160  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA, protein  
A:Residues: 1-851 <F01>  
A>Note: sequence extracted from NCBI backbone (NCBIP:110820)  
R:Yan, R.; Qureshi, S.; Zhong, Z.; Wen, Z.; Darnell, J.E.  
Submitted to the EMBL data library, December 1994  
A:Reference number: S71908  
A:Accession: S71908  
A:Molecule type: DNA  
A:Residues: 1-851 <YAN>  
A:Cross-references: EMBL:U18671; NID:91293919; PIDN:AAA98760.1; PID:91293920  
R:Yan, R.; Qureshi, S.; Zhong, Z.; Wen, Z.; Darnell Jr., J.E.  
Nucleic Acids Res. 23, 459-463, 1995  
A:Title: The genomic structure of the STAT genes: multiple exons in coincident sites in  
A:Reference number: S53873; MUID:95192056; PMID:7885841  
A:Accession: S53873  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-196;392-591;684-730 <YAN>  
A:Cross-references: EMBL:U18671  
C:Genetics:  
A:Gene: stat2  
A:Introns: 44/2; 95/3; 127/3; 157/3; 183/1; 211/3; 261/2; 314/2; 345/2; 365/2; 372/2; 40  
C:Superfamily: human signal transducer and transcription activator STAT3A  
C:Keywords: signal transduction; transcription regulation  
Query Match 1.0%; Score 9; DB 2; Length 851;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 457 ESTLEPVE 465  
DB 756 ESTLEPVE 764  
RESULT 6  
A:0651  
hypothetical protein RV3072C - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: A70651  
R:Coile, S.T.; Broesh, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70651  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-174 <CO>  
A:Cross-references: GB:Z8386; GB:AL123456; NID:93261691; PIDN:CAB06251.1; PID:e290922;  
C:Genetics:  
A:Gene: RV3072C  
Query Match 0.9%; Score 8; DB 2; Length 174;  
Best Local Similarity 100.0%; Pred. No. 9.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 895 RVDVATGV 902

DB 90 RVDVATGV 97  
RESULT 7  
H64989  
membrane-bound tetraheme cytochrome napC - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: H64989  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: H64989  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-200 <BLAT>  
A:Cross-references: GB:A6000309; GB:U00096; NID:91788520; PIDN:AAC75262.1; PID:917885  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: napC  
C:Superfamily: denitrification system component nirT; nirT homology  
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein  
F:18-192/Domain: nirT homology <NIR>  
F:57-60/Binding site: heme (Cys) (covalent) #status predicted  
F:61/Binding site: heme iron (His) (axial ligand) #status predicted  
F:87-90/Binding site: heme (Cys) (covalent) #status predicted  
F:91/Binding site: heme iron (His) (axial ligand) #status predicted  
F:147-150/Binding site: heme (Cys) (covalent) #status predicted  
F:151/Binding site: heme iron (His) (axial ligand) #status predicted  
F:179-182/Binding site: heme (Cys) (covalent) #status predicted  
F:183/Binding site: heme iron (His) (axial ligand) #status predicted  
Query Match 0.9%; Score 8; DB 2; Length 200;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 161 KPGIKRL 168  
DB 7 KPGIKRL 14  
RESULT 8  
C91015  
cytochrome c-type protein [imported] - Escherichia coli (strain O157:H7, substrain RI  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: C91015  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.  
gatawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and 9  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C91015  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-200 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA836514.1; PID:q13362560; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: Ecs3091  
C:Superfamily: denitrification system component nirT; nirT homology  
Query Match 0.9%; Score 8; DB 2; Length 200;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 161 KPGIKRL 168  
DB 7 KPGIKRL 14



979	6	0.7	247	2	T14818	leucine-rich repa
980	6	0.7	247	2	T23763	hypothetical prote
981	6	0.7	247	2	AG3412	DNA repair protei
982	6	0.7	248	1	TVWVRS	transforming prote
983	6	0.7	248	2	G83253	probable short-cha
984	6	0.7	248	2	F64838	hypothetical prote
985	6	0.7	248	2	C90940	hypothetical prote
986	6	0.7	248	2	G85788	hypothetical prote
987	6	0.7	248	2	AC0712	hypothetical prote
988	6	0.7	248	2	S76774	probable outer mem
989	6	0.7	248	2	AD0394	thiol-disulfide in
990	6	0.7	249	2	B83767	hypothetical prote
991	6	0.7	249	2	E69298	glucose 1-dehydrog
992	6	0.7	249	2	U00274	conserved hypothet
993	6	0.7	249	2	AF3325	hypothetical 29k p
994	6	0.7	250	1	OBNC2	hypothetical prote
995	6	0.7	250	1	AB0171	cytochrome-c oxida
996	6	0.7	250	2	E84321	probable 3-deoxy-m
997	6	0.7	250	2	H84312	50S ribosomal prot
998	6	0.7	250	2	C95111	hypothetical prote
999	6	0.7	250	2	F75219	hypothetical prote
1000	6	0.7	250	2	B40626	probable 2,3-dihyd

## ALIGNMENTS

RESULT 1  
 AE0428  
 Probable exported protein YPO3524 [imported] - Yersinia pestis (strain CO92)  
 C:Species: Yersinia pestis  
 C:Date: 02-Nov-2001  
 C:Accession: AE0428  
 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.E.;  
 deno-Farrage, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
 Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:11586360  
 A:Accession: AE0428  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-578 <KUP>  
 A:Cross-references: GB:AL590842; PIDN:G15981448; GSPDB:GM00175  
 C:Genetics:  
 A:Gene: YPO3524

```

Query Match          . 1.4%; Score 13; DB 2; Length 578;
Best Local Similarity 100.0%; Pred. No. 0.00021;
Matches 13; Conservative 0; Mismatches 0; Indels 0;
QY 880 GAGGCGVRNASPYVG 892
    |||
Db 538 GAGGCGVRNASPYVG 550
    |||

```

RESULT 2  
AD1055  
probable exported protein ytfM [imported] - *Salmonella enterica* subsp. *enterica* serovar  
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
A:Note: this species has also been called *Salmonella typhi*  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AD1055  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher  
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar  
S.; Moule, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.  
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* sero  
A:Reference number: AB0502; PMID:11677608  
A:Accession: AD1055  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-577 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD06889.1; PID:g16505537, GSPDB:GN00176  
C:Genetics:  
A:Gene: ytlfM

Query Match	1.2%	Score 11;	DB 2;	Length 577;
Best Local Similarity	100.0%;	Pred. No. 0.023;		
Matches 11; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0

QY	882	GVGVRWASPVG	892
Db	539	GVGVRWASPVG	549

### RESULT 3

C:\Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
C:\Accession: B64012

D.M.: Brandon, R.C.; Fine, L.D.; Frithman, T.; Goeayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodex, A.; Kelley, J.M.; Kirschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.

Science 269, 496-512, 1995  
A: Authors: Gneilm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter

A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: B64012

A;status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-578 <TTCp>

A;Residues: 1-578 <TIGR>  
A;Cross-references: GB:U032752; GB:L42033; NID:q1573692; PIDN:AAC2357.1; PTD:q1573700

Query Match	1.2%	Score 11;	DB 2;	Length 578;
Best Local Similarity	100.0%	Pred. No. 0.024;		
Matches 11; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	882	GVGVRWASPVG	8922
Db	539	GVGVRWASPVG	5499

**RESULT 4**  
**F83327**

C,Species: Pseudomonas aeruginosa  
conserved hypothetical protein PA5543 [imported] -Pseudomonas aeruginosa (C,Species: Pseudomonas aeruginosa  
Date: 15-Sep-2000 #sequence revision 15-Sep-2000 #count altered 31 2 0 00000

C; Accession: F83327  
R; Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;

... Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A, virulence complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Accession*: F83327  
A, reference number: AB2950; NCBI:20437337; PMID:10984043

A;Status: preliminary  
A;Molecule type: DNA

A:Accession: J-015 (S10)  
 A:Cross-references: GB:AE004682; GB:AE004091; NID:g9948598; PIDN:AAG05931.1; GSPDB:GN  
 A:Experimental source: strain PA01

C;Genetics:  
A;Gene: PA2543

Query Match	1.18;	Score 10;	DB 2;	length 579;
Best Local Similarity	100.08;	Pred. No. 0.25;		

809	IRREAGGDS	0;	Indels	0;	Gaps	0;
matches	10;	conservative	0;	Mismatches	0;	

```

Db      470 LRFAGGDQS 479
          |||||

```

## RESULT 5

## DISCUSSION

833	6	0.7	218	2	H64638	conserved hypotet
834	6	0.7	218	2	AH2149	hypothetical prote
835	6	0.7	218	2	B97428	hypothetical prote
836	6	0.7	218	2	G82019	probable periplasm
837	6	0.7	219	2	C75031	lsu ribosomal prot
838	6	0.7	219	2	D71217	probable ribosomal
839	6	0.7	219	2	B98348	probable transcrip
840	6	0.7	219	2	H81824	hypothetical perip
841	6	0.7	219	2	A82498	CbdY family protei
842	6	0.7	219	2	AE0945	probable exported
843	6	0.7	220	1	QOVZC9	F9 protein - sheep
844	6	0.7	220	1	S48551	hypothetical prote
845	6	0.7	220	2	H80329	receptor tyrosine
846	6	0.7	220	2	B72715	hypothetical prote
847	6	0.7	220	2	C72716	hypothetical prote
848	6	0.7	221	2	H83488	probable permease
849	6	0.7	221	2	D91210	probable phosphata
850	6	0.7	221	2	G68056	hypothetical prote
851	6	0.7	221	2	D65174	hypothetical prote
852	6	0.7	221	2	AG0956	probable hydrolase
853	6	0.7	221	2	D90195	hypothetical prote
854	6	0.7	221	2	T52622	probable peptidylp
855	6	0.7	222	1	B69394	probable phosphoe
856	6	0.7	222	2	AH0498	probable haloacid
857	6	0.7	222	2	D75549	sana protein - Dei
858	6	0.7	222	2	H87496	hypothetical prote
859	6	0.7	222	2	E70335	hypothetical prote
860	6	0.7	222	2	S77557	hypothetical prote
861	6	0.7	222	2	B97118	hypothetical prote
862	6	0.7	223	2	S39883	endo-1,4-beta-xyla
863	6	0.7	223	2	H75600	hypothetical prote
864	6	0.7	224	2	D70665	probable urease ac
865	6	0.7	224	2	S65960	maum protein precu
866	6	0.7	224	2	AE3382	alpha/beta hydroa
867	6	0.7	224	2	A64418	hypothetical prote
868	6	0.7	224	2	C91083	hypothetical membr
869	6	0.7	224	2	B75299	sensory transducl
870	6	0.7	224	2	T10660	photosystem II pro
871	6	0.7	224	2	T51875	hypothetical prote
872	6	0.7	225	2	T49769	hypothetical prote
873	6	0.7	225	2	AC1630	cell-division inh
874	6	0.7	225	2	A11267	cell-division inh
875	6	0.7	226	2	T11368	H+-transporting tw
876	6	0.7	226	2	S04752	H+-transporting tw
877	6	0.7	226	2	AF0665	probable amino aci
878	6	0.7	226	2	FA5239	cell-division inh
879	6	0.7	227	2	H75506	oxidoreductase, sh
880	6	0.7	227	2	AG0364	probable two compo
881	6	0.7	227	2	S44836	KO2D10.4 protein -
882	6	0.7	227	2	H90527	glucose inhibited
883	6	0.7	227	2	C83046	probable transcrip
884	6	0.7	228	2	AF0374	probable ABC trans
885	6	0.7	228	2	AF0253	cell division inh
886	6	0.7	228	2	D71359	conserved hypotet
887	6	0.7	228	2	T12799	conserved hypotet
888	6	0.7	228	2	B69777	hypothetical prote
889	6	0.7	228	2	AC2297	hypothetical prote
890	6	0.7	228	2	D90139	conserved hypotet
891	6	0.7	228	2	T48027	hypothetical prote
892	6	0.7	228	2	T45311	cell division alp-
893	6	0.7	229	2	F70800	probable transport
894	6	0.7	229	2	H72318	hypothetical prote
895	6	0.7	229	2	T20848	hypothetical prote
896	6	0.7	229	2	T48636	hypothetical prote
897	6	0.7	229	2	C75301	rRNA methylase Spo
898	6	0.7	230	2	E81397	probable l-acylgly
899	6	0.7	230	2	D81252	hypothetical prote
900	6	0.7	232	2	D84028	septom site-determ
901	6	0.7	232	2	G84382	cobalt transport A
902	6	0.7	232	2	G83609	probable transcrip
903	6	0.7	232	2	E75547	hypothetical prote
904	6	0.7	232	2	E85928	hypothetical prote
905	6	0.7	232	2		
906	6	0.7	233	2	T48179	hypothetical prote
907	6	0.7	233	2	H86925	probable lipoprote
908	6	0.7	233	2	E96903	distantly related
909	6	0.7	234	1	G69024	pharmone shutdown
910	6	0.7	234	2	PC4398	glutamine-tRNA lig
911	6	0.7	234	2	T06995	probable MADS box
912	6	0.7	234	2	T44731	hypothetical prote
913	6	0.7	234	2	F69019	conserved hypotet
914	6	0.7	234	2	A70736	hypothetical prote
915	6	0.7	234	2	D83403	hypothetical prote
916	6	0.7	234	2	A86290	t16N1.11 protein
917	6	0.7	234	2	C75368	conserved hypotet
918	6	0.7	234	2	S60885	ferric exochelin u
919	6	0.7	234	2	AH3335	probable membrane-
920	6	0.7	235	2	A72639	probable uracil ph
921	6	0.7	235	2	G83057	conserved hypotet
922	6	0.7	235	2	B85639	hypothetical prote
923	6	0.7	235	2	B86150	hypothetical prote
924	6	0.7	235	2	AH3210	conserved hypotet
925	6	0.7	236	1	C27058	carboxymethyleneb
926	6	0.7	236	2	S02022	carboxymethyleneb
927	6	0.7	236	2	T44622	carboxymethyleneb
928	6	0.7	236	2	T44669	carboxymethyleneb
929	6	0.7	236	2	S60390	probable membrane
930	6	0.7	236	2	F75375	nodulin 21-related
931	6	0.7	236	2	B86881	hypothetical prote
932	6	0.7	236	2	T50908	hypothetical prote
933	6	0.7	237	2	D71877	probable amino aci
934	6	0.7	237	2	C64637	amino acid ABC tra
935	6	0.7	237	2	T07820	hypothetical prote
936	6	0.7	237	2	AE1757	phase related prot
937	6	0.7	237	2	AE1757	phase protein homo
938	6	0.7	237	2	H83033	probable transcrip
939	6	0.7	237	2	S75270	hypothetical prote
940	6	0.7	238	2	F87605	hypothetical prote
941	6	0.7	238	2	B84951	hypothetical prote
942	6	0.7	238	2	D96026	probable branched-
943	6	0.7	238	2	AH3176	conserved hypotet
944	6	0.7	239	2	T23169	conserved hypotet
945	6	0.7	239	2	B87062	transcription regu
946	6	0.7	239	2	A75585	probable membrane
947	6	0.7	239	2	AE0784	hypothetical prote
948	6	0.7	240	2	G73295	probable ATP-bind
949	6	0.7	240	2	F83488	probable amino aci
950	6	0.7	240	2	H95873	hypothetical prote
951	6	0.7	240	2	H97783	probable ATP-bind
952	6	0.7	240	2	G98155	hypothetical prote
953	6	0.7	240	2	AE3132	deoxynucleotide mo
954	6	0.7	241	1	K1BPDA	ribose-phosphate
955	6	0.7	241	2	F42600	probable octopine
956	6	0.7	241	2	AG3230	hypothetical prote
957	6	0.7	241	2	AI2588	exopolysaccharide
958	6	0.7	241	2	F69150	conserved hypotet
959	6	0.7	241	2	AC2646	conserved hypotet
960	6	0.7	242	2	B83501	hypothetical prote
961	6	0.7	242	2	H83829	hypothetical prote
962	6	0.7	242	2	H70980	hypothetical prote
963	6	0.7	242	2	D97008	probable rsbw prot
964	6	0.7	243	2	G86676	ABC-type polar am
965	6	0.7	243	2	S39400	transcription regu
966	6	0.7	244	1	A75408	hydrogenase (EC 1.
967	6	0.7	244	2	E70544	triosephosphate is
968	6	0.7	244	2	AF2100	phosphatase ABC tr
969	6	0.7	245	2	E70544	1-(5-phosphoribosy
970	6	0.7	245	2	S58176	ketoreductase mmt
971	6	0.7	246	2	S58497	early auxin-induci
972	6	0.7	246	2	G95230	spou RNA Methylas
973	6	0.7	246	2	G72257	hypothetical prote
974	6	0.7	246	2	A87453	hydroxylase, haloac
975	6	0.7	247	2	G87472	hypothetical prote
976	6	0.7	247	2	T22109	hypothetical prote
977	6	0.7	247	2	D84376	proliferating-cell
978	6	0.7	247	2		

687	6	0.7	185	2	C71356	probable ribosomal
688	6	0.7	185	2	H95262	hypothetical prote
689	6	0.7	186	2	D86691	methylinphosphoties
690	6	0.7	186	2	JN0517	lumazine protein L
691	6	0.7	186	2	H86623	polypeptide deform
692	6	0.7	186	2	E72000	conserved hypotet
693	6	0.7	186	2	AD0636	RNA polymerase sig
694	6	0.7	186	2	G87589	hypothetical prote
695	6	0.7	186	2	A75281	gag polyprotein -
696	6	0.7	187	1	FOVW55	glucanokinase (EC
697	6	0.7	187	1	S56494	5-formyltetrahydro
698	6	0.7	187	2	H69956	probable Jpv prot
699	6	0.7	187	2	C70766	hypothetical prote
700	6	0.7	187	2	AE3442	Al4 protein - Indi
701	6	0.7	187	2	S35886	conserved hypotet
702	6	0.7	187	2	A69118	myosin Al catalyti
703	6	0.7	188	2	A23253	myosin light chain
704	6	0.7	189	2	I57590	probable scafold1
705	6	0.7	189	2	AE1446	hypothetical prote
706	6	0.7	189	2	G96514	myosin alkali ligh
707	6	0.7	190	1	MORR1	peptidylprolyl iso
708	6	0.7	190	2	A11001	hypothetical prote
709	6	0.7	190	2	T02619	hypothetical prote
710	6	0.7	190	2	D70828	hypothetical prote
711	6	0.7	190	2	S72571	probable 40S ribos
712	6	0.7	191	2	A96526	ribosomal protein
713	6	0.7	191	2	T07994	recombination prot
714	6	0.7	191	2	C90518	hypothetical prote
715	6	0.7	191	2	H87330	molycoprotein-guan
716	6	0.7	191	2	E75132	hypothetical prote
717	6	0.7	192	2	A75568	transcription regu
718	6	0.7	192	2	T44491	hypothetical prote
719	6	0.7	193	2	T19081	hypothetical prote
720	6	0.7	193	2	F84228	hypothetical prote
721	6	0.7	193	2	C75367	hypothetical prote
722	6	0.7	193	2	T49977	hypothetical prote
723	6	0.7	194	2	C72554	hypothetical prote
724	6	0.7	194	2	A64331	hypothetical prote
725	6	0.7	195	2	S31287	ribosomal protein
726	6	0.7	195	2	A97217	riboserylthrin limpo
727	6	0.7	195	2	S46449	hypothetical prote
728	6	0.7	195	2	AG0778	probable membrane
729	6	0.7	195	2	I50508	achaeete-scute homo
730	6	0.7	196	2	D70467	RNase HII - Aquilfe
731	6	0.7	196	2	E85553	3-methyladenine DN
732	6	0.7	196	2	C72071	DNA-3-methyladenin
733	6	0.7	196	2	G86205	hypothetical prote
734	6	0.7	196	2	JN0721	ATP-corrinoid aden
735	6	0.7	196	2	AB0654	COB(1) alamin aden
736	6	0.7	196	2	D75041	probable scafold1
737	6	0.7	196	2	A11361	protein W03A5.5 [l
738	6	0.7	196	2	G88456	aliphatic amidase
739	6	0.7	196	2	B83226	probable transcrip
740	6	0.7	197	1	T35727	ribosomal protein
741	6	0.7	197	1	S16822	hypothetical prote
742	6	0.7	197	2	E71025	hypothetical prote
743	6	0.7	197	2	F84257	hypothetical prote
744	6	0.7	198	2	T28135	rolp protein - Agr
745	6	0.7	198	2	I39670	thymidine kinase (
746	6	0.7	199	2	D90005	hydrolase family p
747	6	0.7	199	2	C75363	conserved hypotet
748	6	0.7	199	2	A12707	conserved hypotet
749	6	0.7	199	2	AB3497	hypothetical prote
750	6	0.7	200	2	B95057	hypothetical prote
751	6	0.7	200	2	E97926	RNA polymerase (de
752	6	0.7	200	2	S12790	GTP-binding protei
753	6	0.7	200	2	F95985	conserved hypotet
754	6	0.7	200	2	T34654	probable transmemb
755	6	0.7	200	2	E83539	hypothetical prote
756	6	0.7	200	2	T24838	hypothetical prote
757	6	0.7	200	2	AC3623	glutaminase (EC 3.
758	6	0.7	201	2	AH0664	acyl carrier prote
759	6	0.7	201	2	T42647	hypothetical prote
760	6	0.7	201	2	G97247	intracellular prot
761	6	0.7	201	2	S48618	hypothetical prote
762	6	0.7	202	2	G95284	probable WrbA2 tri
763	6	0.7	202	2	B83059	hypothetical prote
764	6	0.7	202	2	A82160	hypothetical prote
765	6	0.7	202	2	E90663	replication prote
766	6	0.7	202	2	A84225	hypothetical prote
767	6	0.7	202	2	A83492	probable transcrip
768	6	0.7	202	2	A85514	replication protei
769	6	0.7	203	2	F82978	guanylate kinase p
770	6	0.7	203	2	S63633	hypothetical prote
771	6	0.7	203	2	T02137	hypothetical prote
772	6	0.7	203	2	F64981	hypothetical 22.4
773	6	0.7	203	2	F91006	hypothetical prote
774	6	0.7	203	2	G85850	hypothetical prote
775	6	0.7	203	2	T08332	conserved hypotet
776	6	0.7	204	2	S76259	second ORF in tran
777	6	0.7	204	2	E70525	hypothetical prote
778	6	0.7	204	2	E82092	hypothetical prote
779	6	0.7	205	2	AC0477	4-methyl-5(B)-hydro
780	6	0.7	205	2	B49910	hemophore Haas [m
781	6	0.7	205	2	E90494	hypothetical prote
782	6	0.7	205	2	H86675	hypothetical prote
783	6	0.7	206	2	AB4988	superoxide dismuta
784	6	0.7	206	2	B82061	30S ribosomal prot
785	6	0.7	206	2	T10113	ribosomal protein
786	6	0.7	206	2	A12674	hypothetical prote
787	6	0.7	206	2	A05159	peptidylprolyl iso
788	6	0.7	206	2	AE0429	hypothetical prote
789	6	0.7	206	2	AF1837	methyalted-DNA-pro
790	6	0.7	207	1	XUHMC	hypothetical prote
791	6	0.7	207	1	B84151	hypothetical prote
792	6	0.7	208	1	A12674	methyalted-DNA-pro
793	6	0.7	209	1	XURTMC	methyalted-DNA-pro
794	6	0.7	209	1	XUHYMC	methyalted-DNA-pro
795	6	0.7	209	2	AT2662	king-kdpg bifunctio
796	6	0.7	209	2	H97444	dhak -type molecule
797	6	0.7	209	2	C44261	probable hydrolase
798	6	0.7	210	2	A81955	glyoxalase II fami
799	6	0.7	210	2	B81011	uridine kinase - D
800	6	0.7	210	2	E75553	probable acetyltra
801	6	0.7	210	2	H75599	dihydrofolate redu
802	6	0.7	211	2	RDBYD	methyalted-DNA-pro
803	6	0.7	211	1	A41809	dolichyl-phosphate
804	6	0.7	211	2	A75176	sodium-type flagel
805	6	0.7	211	2	E82055	hypothetical prote
806	6	0.7	212	2	T32590	dTMP kinase (EC 2.
807	6	0.7	212	2	AH0195	hypothetical prote
808	6	0.7	212	2	F75024	hypothetical prote
809	6	0.7	212	2	A84359	hypothetical prote
810	6	0.7	212	2	T29906	hypothetical prote
811	6	0.7	213	2	F86935	probable hypoxanth
812	6	0.7	213	2	B70707	hypothetical prote
813	6	0.7	213	2	AE3487	probable hypoxanth
814	6	0.7	214	2	I49351	hypothetical prote
815	6	0.7	214	2	UC5043	panthothenate kins
816	6	0.7	214	2	H69339	endothelin 3 precu
817	6	0.7	214	2	C75104	conserved hypotet
818	6	0.7	214	2	C71120	hypothetical prote
819	6	0.7	214	2	AE2934	hypothetical prote
820	6	0.7	214	2	B81247	transcription regu
821	6	0.7	215	2	D75342	EpH/GdmH-related
822	6	0.7	215	2	A70561	conserved hypotet
823	6	0.7	216	2	B97490	probable hpt prote
824	6	0.7	216	2	D89827	lysine decarboxyla
825	6	0.7	217	1	F69098	conserved hypotet
826	6	0.7	217	2	AB6846	phosphate transpor
827	6	0.7	217	2	AE3540	H+-transporting tw
828	6	0.7	217	2	H64419	phosphoglycerate m
829	6	0.7	217	2	B83944	transaldolase (EC
830	6	0.7	217	2	E70574	resolvasse BH2354
831	6	0.7	218	2	S30120	probable transcrip
832	6	0.7	218	2	G71875	hypothetical prote

541	6	0.7	152	2	S56198
542	6	0.7	152	2	A86557
543	6	0.7	152	2	F72066
544	6	0.7	152	2	H81675
545	6	0.7	152	2	H83578
546	6	0.7	153	2	J01226
547	6	0.7	153	2	G82555
548	6	0.7	153	2	D81864
549	6	0.7	153	2	D81079
550	6	0.7	153	2	E84218
551	6	0.7	154	2	T36690
552	6	0.7	154	2	H71517
553	6	0.7	154	2	E81984
554	6	0.7	154	2	A81039
555	6	0.7	154	2	A81951
556	6	0.7	154	2	S66112
557	6	0.7	154	2	AG1021
558	6	0.7	154	2	A99404
559	6	0.7	154	2	G82980
560	6	0.7	154	2	D83516
561	6	0.7	155	2	S38877
562	6	0.7	155	2	E65174
563	6	0.7	156	2	A81015
564	6	0.7	156	2	C72624
565	6	0.7	156	2	I49446
566	6	0.7	156	2	F88175
567	6	0.7	156	2	C84271
568	6	0.7	157	2	B86418
569	6	0.7	158	2	G95186
570	6	0.7	158	2	H98053
571	6	0.7	159	2	H75391
572	6	0.7	159	2	AC2334
573	6	0.7	159	2	B66711
574	6	0.7	159	2	A83219
575	6	0.7	159	2	S72846
576	6	0.7	160	2	G69608
577	6	0.7	160	2	D87481
578	6	0.7	161	2	T39576
579	6	0.7	162	2	A87591
580	6	0.7	162	2	AC0010
581	6	0.7	162	2	B70144
582	6	0.7	163	2	T36985
583	6	0.7	163	2	A75160
584	6	0.7	163	2	D42148
585	6	0.7	164	2	F95026
586	6	0.7	164	2	F97897
587	6	0.7	164	2	C83948
588	6	0.7	164	2	F87438
589	6	0.7	164	2	A97809
590	6	0.7	165	2	H70768
591	6	0.7	165	2	S72810
592	6	0.7	165	2	AH2010
593	6	0.7	166	1	R3B55F
594	6	0.7	166	1	R3B55
595	6	0.7	166	1	S12632
596	6	0.7	166	1	A29240
597	6	0.7	166	1	S12584
598	6	0.7	166	1	T44400
599	6	0.7	166	2	S49101
600	6	0.7	166	2	AB0943
601	6	0.7	166	2	E90471
602	6	0.7	166	2	C33748
603	6	0.7	166	2	S72447
604	6	0.7	167	2	F64953
605	6	0.7	167	2	AH0747
606	6	0.7	167	2	G85803
607	6	0.7	167	2	B90955
608	6	0.7	167	2	AH1975
609	6	0.7	167	2	G86080
610	6	0.7	167	2	G91233
611	6	0.7	167	2	AB1884
612	6	0.7	167	2	AB3204
613	6	0.7	167	2	AB3204

hypothetical prote  
CT406 hypothetical  
conserved hypotet  
conserved hypotet  
H+-transporting tw  
c-type cytochrome  
probable regulator  
regulatory protein  
hypothetical prote  
conserved hypotet  
hypothetical prote  
hypothetical prote  
conserved hypotet  
transcription repr  
probable acetyltra  
hypothetical prote  
conserved hypotet  
hypothetical prote  
tropom C isoform  
tropom C isoform  
hypothetical 16..7  
conserved hypotet  
hypothetical prote  
8-oxo-dGTPase - mo  
protein T24H7.3 [1  
hypothetical prote  
probable plasma me  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
transcription regu  
hypothetical prote  
rRNA methylase hom  
conserved hypotet  
cytochrome c-type  
probable exported  
transcription fact  
hypothetical prote  
3-isopropylmalate  
GTP-binding protei  
ribosomal protein  
30S ribosomal prot  
hypothetical prote  
hypothetical prote  
methylated-DNA-pro  
methylated-DNA-pro  
molybdopterin conv  
ribosomal protein  
ribosomal protein  
cofilin - human  
cofilin - pig  
cofilin - mouse  
ribosomal protein  
extracytoplasmic s  
hypothetical prote  
hypothetical prote  
3-dehydroquinatate  
ferritin-like prot  
ferritin-like prot  
ferritin-like prot  
ferritin-like prot  
ferritin-like prot  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote

614	6	0.7	167	4	S40857
615	6	0.7	168	4	E90403
616	6	0.7	169	2	I64089
617	6	0.7	169	2	D70226
618	6	0.7	170	1	D60010
619	6	0.7	170	2	AE1391
620	6	0.7	170	2	AG1766
621	6	0.7	170	2	B71132
622	6	0.7	170	2	S76067
623	6	0.7	171	1	XUECAD
624	6	0.7	171	2	AF0662
625	6	0.7	171	2	D90868
626	6	0.7	171	2	E85750
627	6	0.7	171	2	F87298
628	6	0.7	171	2	C95143
629	6	0.7	171	2	T20567
630	6	0.7	171	2	A99011
631	6	0.7	172	2	A46695
632	6	0.7	172	2	F87000
633	6	0.7	172	2	H83139
634	6	0.7	172	2	H90430
635	6	0.7	172	2	B86679
636	6	0.7	172	2	H83542
637	6	0.7	173	1	GOCRAA
638	6	0.7	173	1	B45932
639	6	0.7	173	1	GOGABA
640	6	0.7	173	1	D64627
641	6	0.7	173	1	JC5982
642	6	0.7	173	2	S58632
643	6	0.7	173	2	AG3476
644	6	0.7	174	2	E71887
645	6	0.7	174	2	B81145
646	6	0.7	174	2	T03308
647	6	0.7	174	2	F96611
648	6	0.7	174	2	D64150
649	6	0.7	174	2	G69824
650	6	0.7	174	2	A82836
651	6	0.7	174	2	D82629
652	6	0.7	175	2	S03756
653	6	0.7	175	2	PW0051
654	6	0.7	176	2	E69519
655	6	0.7	176	2	A84199
656	6	0.7	176	2	D86434
657	6	0.7	176	2	D87345
658	6	0.7	176	2	C55208
659	6	0.7	177	2	G71360
660	6	0.7	177	2	S45878
661	6	0.7	177	2	T45667
662	6	0.7	178	2	T13213
663	6	0.7	178	2	H83061
664	6	0.7	178	2	F90166
665	6	0.7	178	2	E88637
666	6	0.7	178	2	C84723
667	6	0.7	179	2	D72459
668	6	0.7	179	2	A87654
669	6	0.7	179	2	C87652
670	6	0.7	180	2	C39415
671	6	0.7	180	2	E48909
672	6	0.7	180	2	B69524
673	6	0.7	181	2	F47021
674	6	0.7	181	2	F84934
675	6	0.7	181	2	S56532
676	6	0.7	181	2	T11902
677	6	0.7	182	1	IVMSB
678	6	0.7	182	2	AC2175
679	6	0.7	183	2	E84119
680	6	0.7	183	2	H87707
681	6	0.7	183	2	F90354
682	6	0.7	183	2	I59442
683	6	0.7	184	2	E90335
684	6	0.7	184	2	D72307
685	6	0.7	184	2	S77928
686	6	0.7	185	2	S59560

hypothetical prote  
hypothetical prote  
protein-export pro  
hypothetical prote  
early E1B 20K prot  
H+-transporting AT  
H+-transporting AT  
hypothetical prote  
hypothetical prote  
methylated-DNA-pro  
O6-methylguanine-D  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
conserved hypotet  
epididymal secreto  
probable membrane  
hypothetical prote  
partial transpos  
hypothetical prote  
hypothetical prote  
atrial gland pepti  
atrial gland pepti  
conserved hypotet  
prenylated protein  
hypothetical prote  
methyltransferase  
hypothetical prote  
hypothetical prote  
gene 11 protein, p  
hypothetical prote  
conserved hypotet  
conserved hypotet  
hypothetical prote  
fann protein precu  
synaptojanin 2 del  
conserved hypotet  
hypothetical prote  
protein T17H7.16 [1  
hypothetical prote  
sccA3 protein - My  
methylated-DNA-pro  
hypothetical prote  
major capsid prote  
hypothetical prote  
hypothetical prote  
protein WD9G12.6 [1  
probable PHD-type  
hypothetical prote  
hypothetical prote  
RNA polymerase sig  
fimbrial protein s  
G protein-coupled  
rRNA (adenine-N6')-  
pectic enzyme secr  
transcripton anti  
NADH2 dehydrogenas  
interferon beta pr  
hypothetical prote  
ATP synthase delta  
alkyl hydroperoxid  
receptor tyrosine  
hypothetical prote  
conserved hypotet  
exoskeletal protei  
histone H1.41 - ga

395	6	0.7	107	2	A70966	hypothetical prote
396	6	0.7	107	2	G90307	conserved hypotet
397	6	0.7	108	2	E49898	cellobiose phospho
398	6	0.7	108	2	T13133	protein gp46 - pha
399	6	0.7	108	2	T49144	hypothetical prote
400	6	0.7	108	2	H87647	hypothetical prote
401	6	0.7	109	2	C90768	probable minor tai
402	6	0.7	109	2	E90899	probable minor tai
403	6	0.7	109	2	D90999	probable minor tai
404	6	0.7	109	2	D90997	probable minor tai
405	6	0.7	109	2	C85742	probable tail comp
406	6	0.7	109	2	C85817	hypothetical prote
407	6	0.7	109	2	T42436	FMRFamide-like pep
408	6	0.7	109	2	B72763	hypothetical prote
409	6	0.7	110	2	G65095	hypothetical prote
410	6	0.7	110	2	D91123	hypothetical prote
411	6	0.7	110	2	C85968	probable tRNA synt
412	6	0.7	110	2	S74060	probable tRNA synt
413	6	0.7	110	2	T00622	hypothetical prote
414	6	0.7	111	2	D72730	hypothetical prote
415	6	0.7	112	4	AS9290	hypothetical prote
416	6	0.7	112	4	T10469	hypothetical prote
417	6	0.7	113	2	T01657	usp2 protein - Puc
418	6	0.7	113	2	AD1652	hypothetical prote
419	6	0.7	113	2	AG1587	hypothetical prote
420	6	0.7	113	2	T35806	hypothetical prote
421	6	0.7	113	2	H87313	nitrogen regulator
422	6	0.7	114	2	D84852	hypothetical prote
423	6	0.7	114	2	S65224	probable membrane
424	6	0.7	114	2	T04407	probable membrane
425	6	0.7	115	2	S72762	B1496_C2_194 prote
426	6	0.7	115	2	G90123	40S ribosomal prot
427	6	0.7	115	2	H87573	conserved hypotet
428	6	0.7	116	2	T51015	hypothetical prote
429	6	0.7	116	2	F83226	hypothetical prote
430	6	0.7	116	2	T37070	hypothetical prote
431	6	0.7	116	2	I55064	transposase - Esch
432	6	0.7	118	2	D71671	ribosomal protein
433	6	0.7	119	2	A23063	pancreatic B-cell
434	6	0.7	119	2	H83186	hypothetical prote
435	6	0.7	120	2	AE0892	dihydronoepterin a
436	6	0.7	120	2	F83043	hypothetical prote
437	6	0.7	121	2	T22995	hypothetical prote
438	6	0.7	121	2	B87906	protein F59C6.10 l
439	6	0.7	121	2	R5BS12	ribosomal protein
440	6	0.7	122	1	AC3577	probable regulator
441	6	0.7	123	2	D64003	probable sodium-tr
442	6	0.7	123	2	AG5136	mutator mult prote
443	6	0.7	124	2	C69478	conserved hypotet
444	6	0.7	124	2	AC0803	NADH dehydrogenase
445	6	0.7	124	2	E90350	probable bacteriop
446	6	0.7	124	2	AB0928	hypothetical prote
447	6	0.7	125	2	F87404	probable membrane
448	6	0.7	125	2	H69171	ribosomal protein
449	6	0.7	127	2	AS4670	hypothetical prote
450	6	0.7	127	2	A28084	RNA polymerase II
451	6	0.7	127	2	A70502	DNA-binding protei
452	6	0.7	127	2	PC2260	probable lprj prot
453	6	0.7	128	2	AB0070	cytochrome P450 pr
454	6	0.7	128	2	B69779	mutator protein Mu
455	6	0.7	128	2	T48606	conserved hypotet
456	6	0.7	129	2	AD3347	hypothetical prote
457	6	0.7	129	2	S61838	LSU ribosomal prot
458	6	0.7	129	2	D86487	che22 protein - Rh
459	6	0.7	130	2	E69384	14.3K hypothetical
460	6	0.7	131	1	H5ZPA3	conserved hypotet
461	6	0.7	131	1	IIECA2	histone H2A.2 - fl
462	6	0.7	131	2	AB0444	hypothetical prote
463	6	0.7	131	2	H75573	4-carboxymuconolac
464	6	0.7	131	2	AB7450	hypothetical prote
465	6	0.7	131	2		
466	6	0.7	131	2		
467	6	0.7	131	2		

histone H2A.1 - fl  
 hypothetical prote  
 hypothetical prote  
 hypothetical prote  
 competence-damage  
 hypothetical prote  
 minor capsid prote  
 Ild894 (Af322013) f  
 hypothetical prote  
 natruiretic peptid  
 probable transcrip  
 conserved hypotet  
 pol protein - ovin  
 methylmalonyl-CoA  
 sigma-B regulator  
 hypothetical prote  
 hypothetical prote  
 transcription term  
 transcription term  
 transcription term  
 beta-1,3-glucanase  
 protein gp45 - pha  
 conserved hypotet  
 hypothetical prote  
 hypothetical prote  
 thiooredoxin - Dein  
 hemoglobin alpha-A  
 probable regulator  
 protein C, osmotic  
 ync homolog Ape02  
 osmotically induci  
 osmotically induci  
 osmotically induci  
 hypothetical prote  
 Ig mu heavy chain  
 conserved hypotet  
 hypothetical prote  
 hypothetical prote  
 probable acetyltra  
 ribosomal-protein-  
 protein T32E20.11  
 hypothetical prote  
 hypothetical prote  
 snRNP core protein  
 phosphotransferase  
 probable glycine C  
 hypothetical prote  
 hypothetical prote  
 hypothetical prote  
 mult like protein  
 40S ribosomal prot  
 hypothetical prote  
 conserved hypotet  
 hypothetical prote  
 hypothetical prote  
 hypothetical prote  
 ccl2 protein - Rho  
 LSU ribosomal prot  
 hypothetical prote  
 hypothetical prote  
 myosin alkali 11gh  
 superoxide dismuta  
 myosin A2 catalyti  
 hypothetical prote  
 hypothetical prote  
 sepd protein - Esc  
 transcription regu

249	7	0.8	875	2	D81651
250	7	0.8	947	2	E86362
251	7	0.8	949	2	D90803
252	7	0.8	987	2	T40241
253	7	0.8	992	2	T38817
254	7	0.8	993	2	A10669
255	7	0.8	996	2	E98200
256	7	0.8	996	2	AD3086
257	7	0.8	1005	2	H85611
258	7	0.8	1026	2	G81751
259	7	0.8	1068	1	A43322
260	7	0.8	1068	1	T38110
261	7	0.8	1134	2	S53955
262	7	0.8	1157	2	F97255
263	7	0.8	1223	2	S62011
264	7	0.8	1228	2	T40468
265	7	0.8	1252	2	S36016
266	7	0.8	1259	2	A43425
267	7	0.8	1268	1	A39640
268	7	0.8	1286	1	T02187
269	7	0.8	1292	2	T48007
270	7	0.8	1300	2	I53799
271	7	0.8	1302	2	C81182
272	7	0.8	1322	2	T01842
273	7	0.8	1335	2	S07245
274	7	0.8	1342	2	A31946
275	7	0.8	1356	2	S32763
276	7	0.8	1388	2	A53317
277	7	0.8	1434	2	G71232
278	7	0.8	1475	2	F63399
279	7	0.8	1539	2	T30037
280	7	0.8	1829	2	S35027
281	7	0.8	1829	2	E81066
282	7	0.8	1870	2	S37671
283	7	0.8	1872	2	S36152
284	7	0.8	1886	2	S04921
285	7	0.8	1914	2	T42635
286	7	0.8	1939	2	AF0095
287	7	0.8	2142	2	B35098
288	7	0.8	2336	2	A45386
289	7	0.8	2698	2	B96671
290	7	0.8	3157	2	B70969
291	7	0.8	3415	2	A46105
292	7	0.8	3716	2	E70969
293	7	0.8	7829	2	T15789
294	6	0.7	12	2	S29830
295	6	0.7	22	2	S05236
296	6	0.7	25	2	T09385
297	6	0.7	32	2	T17394
298	6	0.7	33	2	A56818
299	6	0.7	42	2	S7815
300	6	0.7	51	2	S01870
301	6	0.7	52	2	S44790
302	6	0.7	52	2	G82835
303	6	0.7	55	2	B98004
304	6	0.7	55	2	AH2396
305	6	0.7	56	2	A29235
306	6	0.7	57	2	C84213
307	6	0.7	60	2	T03816
308	6	0.7	61	2	S60796
309	6	0.7	61	2	D69526
310	6	0.7	62	2	PX0018
311	6	0.7	63	2	C96942
312	6	0.7	64	2	B82672
313	6	0.7	66	2	S77700
314	6	0.7	66	2	H86090
315	6	0.7	68	2	AF0066
316	6	0.7	69	2	S70158
317	6	0.7	69	2	F90926
318	6	0.7	69	2	B85775
319	6	0.7	69	2	C64925
320	6	0.7	70	2	AH0700
321	6	0.7	71	2	T50483

conserved hypotet  
hypothetical prote  
Aida-1 adhesin-lik  
probable guanin n  
hypothetical prote  
probable virulence  
sarcosine oxidase  
probable adhesin Z  
exodeoxyribonuclea  
1-phosphatidylinos  
hypothetical prote  
fusion of alpha-g1  
PHO85 protein - ye  
surface layer prot  
ocyst wall protei  
Bravo/Nr-CAM cell  
neural cell adhesi  
probable ABC trans  
P-glycoprotein hom  
CG1 protein - huma  
iron-regulated pro  
hypothetical prote  
xanthine dehydroge  
xanthine dehydroge  
kinecin 1 - human  
collagen alpha 1(X  
hypothetical prote  
protein FI1L21.22  
hypothetical prote  
cytotoxin RTX homo  
iron-regulated pro  
MHC class III hist  
MHC class III hist  
nuclear pore prote  
tenascin Y precus  
probable sideropho  
MHC class III hist  
omega-conotoxin-se  
similar to transla  
probable PPE prote  
polyprotein(NS1, N  
probable PPE prote  
hypothetical prote  
dimethylalliline mo  
exoenzyme C3 - Clo  
rev protein - huma  
viral protein - Dic  
Na+/K+-exchanging  
probable DNA ligas  
hypothetical prote  
D2007.3 protein -  
hypothetical prote  
hypothetical prote  
hemolymph trypsin  
30S ribosomal prot  
hypothetical prote  
M protein precurs  
conserved hypotet  
GTP-binding protel  
hypothetical prote  
thiol protein - Es  
thiamin biosynthes  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
conserved hypotet  
glucose-repressibl

322	6	0.7	72	2	E97706
323	6	0.7	73	2	T31216
324	6	0.7	74	2	S28794
325	6	0.7	74	2	S63632
326	6	0.7	74	2	B87294
327	6	0.7	74	2	G85899
328	6	0.7	76	2	S26943
329	6	0.7	76	2	B96809
330	6	0.7	77	2	JH0649
331	6	0.7	78	1	G69508
332	6	0.7	78	2	T03670
333	6	0.7	78	2	S61473
334	6	0.7	78	2	S61475
335	6	0.7	78	2	S61474
336	6	0.7	79	2	C87028
337	6	0.7	79	2	H70827
338	6	0.7	79	2	AC0013
339	6	0.7	81	2	T03723
340	6	0.7	81	2	T03707
341	6	0.7	81	2	AC0142
342	6	0.7	82	2	H83584
343	6	0.7	82	2	AF3441
344	6	0.7	83	2	G87393
345	6	0.7	86	2	C81052
346	6	0.7	86	2	S60073
347	6	0.7	86	2	C97450
348	6	0.7	86	2	AE2668
349	6	0.7	87	2	AF0910
350	6	0.7	87	2	T47209
351	6	0.7	88	2	S60188
352	6	0.7	88	2	C83657
353	6	0.7	88	2	AG1356
354	6	0.7	88	2	AH1726
355	6	0.7	89	2	A84301
356	6	0.7	89	2	T28989
357	6	0.7	90	2	T46580
358	6	0.7	91	2	T05920
359	6	0.7	91	2	S52275
360	6	0.7	91	2	D75561
361	6	0.7	91	2	H97856
362	6	0.7	92	2	A84101
363	6	0.7	93	2	B90684
364	6	0.7	93	2	S37677
365	6	0.7	93	2	F85534
366	6	0.7	93	2	G64886
367	6	0.7	93	2	H64767
368	6	0.7	94	2	B72867
369	6	0.7	94	2	S75465
370	6	0.7	94	2	S32939
371	6	0.7	95	1	L0F610
372	6	0.7	96	1	FEAH
373	6	0.7	96	2	AB3053
374	6	0.7	97	2	I51262
375	6	0.7	97	2	JC1139
376	6	0.7	97	2	B28489
377	6	0.7	98	2	C86517
378	6	0.7	98	2	G84296
379	6	0.7	98	2	D72106
380	6	0.7	100	2	A32282
381	6	0.7	100	2	F72530
382	6	0.7	102	2	I53094
383	6	0.7	102	2	AI2693
384	6	0.7	102	2	G89574
385	6	0.7	104	2	G89574
386	6	0.7	105	1	RGECMU
387	6	0.7	105	2	C91237
388	6	0.7	105	2	G86084
389	6	0.7	105	2	T44766
390	6	0.7	105	2	G89993
391	6	0.7	106	2	T36932
392	6	0.7	106	2	AG0216
393	6	0.7	106	2	F87608
394	6	0.7	106	2	AI2219

hypothetical prote  
hypothetical prote  
H+-transporting tw  
H+-transporting tw  
AMP synthase F0, C  
probable oxidoredu  
H+-transporting tw  
protein F28K19.18  
rab protein Rab19  
conserved hypotet  
reverse transcript  
p83/100 protein -  
p83/100 protein -  
p83/100 protein -  
hypothetical prote  
hypothetical prote  
hypothetical prote  
reverse transcript  
reverse transcript  
conserved hypotet  
probable biotin-re  
hypothetical prote  
acyl carrier prote  
veg protein - Bact  
exodeoxyribonuclea  
conserved hypotet  
ccg-4 protein 2 (1  
11beta-hydroxyster  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
2'-aminobiphenyl-2  
probable cysteine  
hypothetical prote  
Mutr/nudix family  
hypothetical prote  
hypothetical prote  
hypothetical prote  
unknown [imported]  
probable membrane  
hypothetical prote  
fibrous body prote  
hypothetical prote  
AVAL protein - yea  
calpactin I light  
ferredoxin (2Fe-2S  
hypothetical prote  
aggreacan chondroit  
calpactin I light  
calpactin I light  
hypothetical prote  
hypothetical prote  
Alzheimer's diseas  
hypothetical prote  
gene Pax-1 protein  
conserved hypotet  
conserved hypotet  
Met regulon regula  
repressor of all m  
transferrin complex p  
hypothetical prote  
probable integral  
conserved hypotet  
hypothetical prote  
hypothetical prote

103	7	0.8	323	2	AH0903	octaprenyl-diphosp
104	7	0.8	323	2	AC2113	hypothetical prote
105	7	0.8	331	2	T02349	glucan endo-1,3-be
106	7	0.8	337	2	A12360	hypothetical prote
107	7	0.8	338	2	AS4078	probable membrane
108	7	0.8	340	2	A12922	hypothetical prote
109	7	0.8	341	2	C89789	hypothetical prote
110	7	0.8	343	2	D83282	probable permease
111	7	0.8	344	2	B97697	sugar ABC transpor
112	7	0.8	347	2	A65022	hypothetical prote
113	7	0.8	347	2	H91044	hypothetical prote
114	7	0.8	347	2	C85889	hypothetical prote
115	7	0.8	351	2	T44428	probable gonococci
116	7	0.8	352	2	E82263	conserved hypotbet
117	7	0.8	355	2	D84192	glutamate decarbox
118	7	0.8	356	2	T00881	probable PCF2-like
119	7	0.8	358	2	H87267	heat-inducible tra
120	7	0.8	368	2	T36414	probable iron-side
121	7	0.8	368	2	AC1149	homoserine O-acety
122	7	0.8	368	2	AC1508	homoserine O-acety
123	7	0.8	371	2	B83781	two-component sens
124	7	0.8	375	1	A23686	limulus clotting e
125	7	0.8	379	2	F72022	hypothetical prote
126	7	0.8	379	2	H86600	hypothetical prote
127	7	0.8	382	2	B82177	Na+/H+ antiporter
128	7	0.8	382	2	D91038	hypothetical prote
129	7	0.8	382	2	G91276	hypothetical prote
130	7	0.8	383	2	JY0360	Na+/H+-exchanging
131	7	0.8	383	2	A70612	hypothetical prote
132	7	0.8	385	2	A83466	probable RND efflu
133	7	0.8	389	2	T22465	hypothetical prote
134	7	0.8	389	2	H82825	hypothetical prote
135	7	0.8	395	2	S74950	transport protein
136	7	0.8	396	2	G69808	patat protein - Sy
137	7	0.8	397	2	T35609	multidrug resistan
138	7	0.8	398	2	A37274	while protein I - S
139	7	0.8	401	2	AC2113	glia-derived nexin
140	7	0.8	402	2	C64895	alanine racemase I
141	7	0.8	402	2	G85882	hypothetical prote
142	7	0.8	402	2	G85882	probable virulence
143	7	0.8	402	2	G86117	probable virulence
144	7	0.8	402	2	AH0514	probable IS elemen
145	7	0.8	402	2	T25001	hypothetical prote
146	7	0.8	405	2	A35620	coenzyme F420 hydr
147	7	0.8	426	2	F69723	trigger factor tig
148	7	0.8	426	2	D97566	deoxyguanosinetrip
149	7	0.8	426	2	AB2787	cytochrome P450 Rv
150	7	0.8	428	1	F70729	hypothetical prote
151	7	0.8	430	2	T46099	hypothetical prote
152	7	0.8	432	2	AF0567	probable permease
153	7	0.8	435	2	G90700	probable uracil tr
154	7	0.8	435	2	B85551	probable transport
155	7	0.8	435	2	H64782	probable membrane
156	7	0.8	441	2	D95390	hypothetical prote
157	7	0.8	451	2	B96495	hypothetical prote
158	7	0.8	454	2	E70637	hypothetical prote
159	7	0.8	465	2	C70594	probable phosphom
160	7	0.8	471	2	E87340	M20/M25/M40 family
161	7	0.8	473	2	T46049	RuvB DNA helicase-
162	7	0.8	477	2	C81865	succinate-semialde
163	7	0.8	480	2	E83105	exodeoxyribonuclea
164	7	0.8	482	2	E70554	hypothetical prote
165	7	0.8	488	2	S38987	hypothetical prote
166	7	0.8	496	2	T50272	alpha, alpha-trehal
167	7	0.8	496	2	S50518	probable aldehyde
168	7	0.8	496	2	A56040	hypothetical prote
169	7	0.8	499	2	A12449	protein-tyrosine k
170	7	0.8	505	2	C96300	hypothetical prote
171	7	0.8	507	2	AD2983	hypothetical prote
172	7	0.8	512	1	C53308	probable hydrolase
173	7	0.8	512	1	ALBSL	alpha-amyase (EC
174	7	0.8	512	2	G91200	type III secretion
175	7	0.8	512	2	T80311	sepc protein - Esc
175	7	0.8	512	2	C86047	esc [imported] -
176	7	0.8	513	2	S54469	hypothetical prote
177	7	0.8	514	2	A87266	sensor histidine k
178	7	0.8	516	2	F86327	protein F18014.21
179	7	0.8	518	2	AB3566	type I restriction
180	7	0.8	519	1	A54549	glucan 1,4-alpha-g
181	7	0.8	519	2	B45697	immediate-early pr
182	7	0.8	521	2	S02018	regulatory protein
183	7	0.8	522	2	E82024	peptide methionine
184	7	0.8	522	2	G81243	probable spectridin
185	7	0.8	523	2	H70886	probable PE protei
186	7	0.8	528	2	A70915	hypothetical prote
187	7	0.8	533	2	D71202	probable ATP-bindi
188	7	0.8	537	2	A64085	hypothetical prote
189	7	0.8	552	1	S76051	hypothetical prote
190	7	0.8	559	2	B72487	probable cytochrom
191	7	0.8	566	2	D84611	methy1-accepting c
192	7	0.8	566	2	A72254	Ras pathway intera
193	7	0.8	567	2	T43535	conserved hypotbet
194	7	0.8	567	2	D83400	probable DNA bindi
195	7	0.8	567	2	AB0006	hypothetical 64.6K
196	7	0.8	571	2	T39378	sensor histidine k
197	7	0.8	574	2	E65135	cerebroglycan prec
198	7	0.8	574	2	B87619	hypothetical prote
199	7	0.8	579	2	A9649	glucose-6-phosphat
200	7	0.8	584	2	B90281	inner membrane cop
201	7	0.8	584	2	T14631	mbp-related protei
202	7	0.8	587	2	F83336	adenine deaminase
203	7	0.8	595	1	A48077	adenine deaminase
204	7	0.8	597	2	D98186	beta-galactosidase
205	7	0.8	597	2	D98186	delhydroxy-acid deh
206	7	0.8	612	2	E82756	exonuclease ABC,
207	7	0.8	614	2	A95072	conserved hypotbet
208	7	0.8	614	2	G97939	outer membrane pro
209	7	0.8	617	2	G82706	hypothetical prote
210	7	0.8	617	2	A13488	numb-binding prote
211	7	0.8	623	2	T04562	Rppp-related prot
212	7	0.8	627	2	T09458	hypothetical prote
213	7	0.8	628	2	G72499	unknown protein F1
214	7	0.8	629	2	T48799	adapler protein CM
215	7	0.8	629	2	T34370	acyl-CoA oxidase h
216	7	0.8	630	2	G96656	hypothetical prote
217	7	0.8	639	2	T13151	conserved hypotbet
218	7	0.8	641	2	T44585	transporter BME107
219	7	0.8	652	2	B83725	DNA ligase - Therm
220	7	0.8	679	2	D69329	synthaxin-binding p
221	7	0.8	682	2	AG3351	probable transcrip
222	7	0.8	688	2	D72418	hypothetical prote
223	7	0.8	693	2	T41443	hypothetical prote
224	7	0.8	695	2	B95879	numb-binding prote
225	7	0.8	698	2	A69222	conserved secreted
226	7	0.8	706	2	G72499	replication licens
227	7	0.8	728	2	T09457	hypothetical prote
228	7	0.8	732	2	AB2732	hypothetical prote
229	7	0.8	741	2	B97513	probable secreted
230	7	0.8	744	1	IS1022	replication licens
231	7	0.8	745	2	T03119	hypothetical prote
232	7	0.8	752	2	G82798	hypothetical prote
233	7	0.8	767	2	T05662	hypothetical prote
234	7	0.8	776	2	T15411	hypothetical prote
235	7	0.8	795	1	S22804	replication licens
236	7	0.8	804	2	H75549	glycogen debranchl
237	7	0.8	807	1	IS1685	replication licens
238	7	0.8	808	1	S62594	replication licens
239	7	0.8	815	2	T00538	probable serine pr
240	7	0.8	817	2	T16409	hypothetical prote
241	7	0.8	830	2	F83288	vacuolar membrane
242	7	0.8	830	2	S25198	hypothetical prote
243	7	0.8	839	2	T45908	probable virulence
244	7	0.8	846	2	AD0279	hypothetical trans
245	7	0.8	850	2	AF3484	probable aminopept
246	7	0.8	862	2	G87094	probable trehalose
247	7	0.8	862	2	E84567	hypothetical prote
248	7	0.8	862	2	T01141	hypothetical prote

GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: April 28, 2003, 16:32:03 ; Search time 31 Seconds  
(without alignments)  
2849,920 Million cell updates/sec

Title: US-09-914-168-2  
Perfect score: 919  
Sequence: 1 MSKPVLPANRSEMPVALAA.....TGKKEGCKPIKLHFICTPF 919

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :  
1: PIR73:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	1.4	578	2	AE0428
2	11	1.2	577	2	AD0455
3	11	1.2	578	2	B64012
4	10	1.1	579	2	F83327
5	9	1.0	851	2	A46160
6	8	0.9	174	2	A70651
7	8	0.9	200	2	H64989
8	8	0.9	200	2	C91015
9	8	0.9	200	2	E85859
10	8	0.9	331	2	AE1390
11	8	0.9	331	2	AG1765
12	8	0.9	340	2	T44739
13	8	0.9	360	2	T27569
14	8	0.9	380	2	JC5747
15	8	0.9	469	2	T13465
16	8	0.9	577	2	S36445
17	8	0.9	577	2	F91278
18	8	0.9	577	2	F86119
19	8	0.9	582	2	F82064
20	8	0.9	731	2	AH2546
21	8	0.9	1355	2	S40022
22	8	0.9	1402	2	S42748
23	8	0.9	1417	2	T00661
24	7	0.8	66	2	G82667
25	7	0.8	72	2	S75519
26	7	0.8	82	2	G55116
27	7	0.8	82	2	AG0884
28	7	0.8	82	2	C85958
29	7	0.8	82	2	C91113

30	7	0.8	88	2	AH0088	probable flagellar
31	7	0.8	91	2	A81784	hypothetical prote
32	7	0.8	95	2	T18004	hypothetical prote
33	7	0.8	107	2	S61135	hypothetical prote
34	7	0.8	109	2	S73639	M635 homolog G12-
35	7	0.8	119	1	A69441	conserved hypotet
36	7	0.8	119	2	A97825	50S ribosomal prot
37	7	0.8	124	2	A48075	60K stress-related
38	7	0.8	129	2	F97593	50S ribosomal prot
39	7	0.8	129	2	AC2815	hypothetical prote
40	7	0.8	130	2	AC1545	hypothetical prote
41	7	0.8	130	2	AD1187	hypothetical prote
42	7	0.8	136	2	CG9892	Dnak suppressor pr
43	7	0.8	138	2	F81243	probable Dnak supp
44	7	0.8	138	2	G82015	hypothetical prote
45	7	0.8	144	2	F72556	hypothetical prote
46	7	0.8	157	2	A84402	hypothetical prote
47	7	0.8	159	2	B70635	conserved hypotet
48	7	0.8	164	2	G72262	probable l(+)-tart
49	7	0.8	166	2	D71049	hypothetical prote
50	7	0.8	174	2	CG4174	hypothetical prote
51	7	0.8	177	2	F96739	hypothetical prote
52	7	0.8	177	2	E83545	ferric pseudobacti
53	7	0.8	179	2	S59503	23k calcium-bindin
54	7	0.8	190	2	A33839	hypothetical prote
55	7	0.8	194	2	C96740	interferon alpha-I
56	7	0.8	195	1	IVBO11	hypothetical prote
57	7	0.8	202	2	F82590	hypothetical prote
58	7	0.8	206	2	G97061	formate dehydrogen
59	7	0.8	211	1	A40836	formate dehydrogen
60	7	0.8	211	2	A86078	formate dehydrogen
61	7	0.8	211	2	B91231	formate dehydrogen
62	7	0.8	211	2	AD0946	fix protein
63	7	0.8	211	2	S15167	conserved hypotet
64	7	0.8	212	2	G75458	hypothetical prote
65	7	0.8	213	2	A95013	hypothetical prote
66	7	0.8	213	2	D97884	hypothetical prote
67	7	0.8	214	2	T21585	hypothetical prote
68	7	0.8	214	2	G81237	hypothetical prote
69	7	0.8	225	2	C95082	amino acid ABC tra
70	7	0.8	225	2	G97949	phosphoglycolate p
71	7	0.8	226	2	F82052	hypothetical prote
72	7	0.8	231	2	C83015	hypothetical prote
73	7	0.8	232	2	S26023	cytochrome-c oxida
74	7	0.8	239	2	H82010	hypothetical prote
75	7	0.8	242	2	B64917	transcription regu
76	7	0.8	242	2	B90918	hypothetical prote
77	7	0.8	242	2	G85766	hypothetical prote
78	7	0.8	243	2	AC0690	probable two-compo
79	7	0.8	245	2	AB0499	ABC transporter pe
80	7	0.8	251	2	C72768	probable electon
81	7	0.8	256	2	AG2545	hypothetical prote
82	7	0.8	266	2	F70595	hypothetical prote
83	7	0.8	271	2	C86907	regulator of purin
84	7	0.8	277	1	F69398	dihydroxyphenolic
85	7	0.8	280	2	A49921	conserved hypotet
86	7	0.8	281	2	A43749	lignin beta-ether
87	7	0.8	282	2	E93554	hypothetical prote
88	7	0.8	285	2	H63659	branched-chain ami
89	7	0.8	293	2	C71267	probable carboxyle
90	7	0.8	294	2	CG4048	major ferric iron-
91	7	0.8	297	2	JC4398	thiosulfate sulfur
92	7	0.8	297	2	JC5286	thiosulfate sulfur
93	7	0.8	298	2	CG1516	hypothetical prote
94	7	0.8	303	2	H70716	hypothetical prote
95	7	0.8	305	2	C97405	conserved hypotet
96	7	0.8	310	2	AC2623	hypothetical prote
97	7	0.8	310	2	F86737	hypothetical prote
98	7	0.8	313	2	AG1656	glycerate dehydrog
99	7	0.8	317	2	S51572	moca prenyl - Rhl
100	7	0.8	323	2	E65109	octaprenyl-diphosp
101	7	0.8	323	2	B91137	octaprenyl diphosp
102	7	0.8	323	2	E85982	octaprenyl diphosp



```

; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C31
; CURRENT APPLICATION NUMBER: US/10/174,579
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-579-356

```

```

Query Match          0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 404 LOAVRAL 410
Db 76 LOAVRAL 82

```

```

RESULT 40
US-10-174-582-356
; Sequence 356, Application US/10174582
; Publication No. US20030027265A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C36
; CURRENT APPLICATION NUMBER: US/10/174,582
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-582-356

```

```

Query Match          0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 404 LOAVRAL 410
Db 76 LOAVRAL 82

```

Search completed: April 28, 2003, 16:41:50  
 Job time : 88 secs

```

; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C13
; CURRENT APPLICATION NUMBER: US/10/180,552
; CURRENT FILING DATE: 2002-06-25
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-180-552-356

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 404 LQAVRAL 410
DB 76 LQAVRAL 82

RESULT 36
US-10-180-557-356
; Sequence 356, Application US/10180557
; Publication No. US20030022301A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C147
; CURRENT APPLICATION NUMBER: US/10/180,557
; CURRENT FILING DATE: 2002-06-25
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-180-557-356

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 404 LQAVRAL 410
DB 76 LQAVRAL 82

RESULT 37
US-10-173-700-356
; Sequence 356, Application US/10173700
; Publication No. US20030027262A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
```

```

; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C14
; CURRENT APPLICATION NUMBER: US/10/173,700
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-173-700-356

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 404 LQAVRAL 410
DB 76 LQAVRAL 82

RESULT 38
US-10-174-572-356
; Sequence 356, Application US/10174572
; Publication No. US20030027263A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C40
; CURRENT APPLICATION NUMBER: US/10/174,572
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-174-572-356

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 404 LQAVRAL 410
DB 76 LQAVRAL 82

RESULT 39
US-10-174-579-356
; Sequence 356, Application US/10174579
; Publication No. US20030027264A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
```

```

; FILE REFERENCE: P3430RIC60
; CURRENT APPLICATION NUMBER: US/10/175,752
; CURRENT FILING DATE: 2002-06-19
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC60
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-752-356

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410
DB 76 LQAVRAL 82

RESULT 32
US-10-176-482-356
; Sequence 356, Application US/10176482
; Publication No. US20030022296A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC60
; CURRENT APPLICATION NUMBER: US/10/176,482
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-482-356

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410
DB 76 LQAVRAL 82

RESULT 33
US-10-176-757-356
; Sequence 356, Application US/10176757
; Publication No. US20030022297A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC60
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-757-356

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410
DB 76 LQAVRAL 82
```

```

; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC60
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-757-356

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410
DB 76 LQAVRAL 82

RESULT 34
US-10-176-913-356
; Sequence 356, Application US/10176913
; Publication No. US20030022298A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Matanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC60
; CURRENT APPLICATION NUMBER: US/10/176,913
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-913-356

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410
DB 76 LQAVRAL 82

RESULT 35
US-10-180-552-356
; Sequence 356, Application US/10180552
; Publication No. US20030022300A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC60
; CURRENT APPLICATION NUMBER: US/10/180,552
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-180-552-356

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410
DB 76 LQAVRAL 82
```

```
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-356
```

```
Query Match          0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 404 LQAVRAL 410
      |||||
Db 76 LQAVRAL 82
```

## RESULT 28

```
US-10-063-502-118
; Sequence 118, Application US/10063502
; Publication No. US20030023042A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,502
; CURRENT FILING DATE: 2002-05-01
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 118
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-502-118
```

```
Query Match          0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 404 LQAVRAL 410
      |||||
Db 76 LQAVRAL 82
```

## RESULT 29

```
US-10-173-706-356
; Sequence 356, Application US/10173706
; Publication No. US20030022293A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C7
; CURRENT APPLICATION NUMBER: US/10/173,706
; CURRENT FILING DATE: 2002-06-17
```

```
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-706-356
```

```
Query Match          0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 404 LQAVRAL 410
      |||||
Db 76 LQAVRAL 82
```

## RESULT 30

```
US-10-175-738-356
; Sequence 356, Application US/10175738
; Publication No. US20030022294A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C45
; CURRENT APPLICATION NUMBER: US/10/175,738
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 356
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-738-356
```

```
Query Match          0.8%; Score 7; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 404 LQAVRAL 410
      |||||
Db 76 LQAVRAL 82
```

## RESULT 31

```
US-10-175-752-356
; Sequence 356, Application US/10175752
; Publication No. US20030022295A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
```

; ORGANISM: Homo Sapien  
US-10-063-547-118

Query Match  
Best Local Similarity 0.8%; Score 7; DB 9; Length 261;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410  
Db 76 LQAVRAL 82

## RESULT 24

US-10-174-590-356  
; Sequence 356, Application US/10174590  
; Publication No. US20030008352A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C42  
; CURRENT APPLICATION NUMBER: US/10/174.590  
; CURRENT FILING DATE: 2002-06-18  
; Prior application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 356  
; LENGTH: 261  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-174-590-356

Query Match  
Best Local Similarity 0.8%; Score 7; DB 9; Length 261;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410  
Db 76 LQAVRAL 82

## RESULT 25

US-10-176-758-356  
; Sequence 356, Application US/10176758  
; Publication No. US20030008353A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C104  
; CURRENT APPLICATION NUMBER: US/10/176.758  
; CURRENT FILING DATE: 2002-06-21  
; Prior application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 356  
; LENGTH: 261  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-176-758-356

Query Match  
Best Local Similarity 0.8%; Score 7; DB 9; Length 261;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410  
Db 76 LQAVRAL 82

## RESULT 26

US-10-063-616-118  
; Sequence 118, Application US/10063616  
; Publication No. US20030013855A1  
; GENERAL INFORMATION:  
; APPLICANT: Baton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/063.616  
; CURRENT FILING DATE: 2002-05-03  
; Prior application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 118  
; LENGTH: 261  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-063-616-118

Query Match  
Best Local Similarity 0.8%; Score 7; DB 9; Length 261;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL 410  
Db 76 LQAVRAL 82

## RESULT 27

US-10-175-737-356  
; Sequence 356, Application US/10175737  
; Publication No. US20030013153A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C50  
; CURRENT APPLICATION NUMBER: US/10/175.737  
; CURRENT FILING DATE: 2002-06-19  
; Prior application removed - See File Wrapper or Palm

```
; SEQ ID NO 162
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Glycine max
US-10-062-254-162
```

```
Query Match          0.8%; Score 7; DB 12; Length 188;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 393 LITVNMG 399
|||||
DB 46 LITVNMG 52
```

```
RESULT 20
US-09-764-868-956
; Sequence 956, Application US/09764868
; Patent No. US2002016871A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIORITY FILING DATE: 2001-01-17
; Prior Application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 956
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
; NAME/KEY: SITE
; LOCATION: (183)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (204)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (209)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-764-868-956
```

```
Query Match          0.8%; Score 7; DB 9; Length 214;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 508 DRYLAIN 514
|||||
DB 106 DRYLAIN 112
```

```
RESULT 21
US-09-825-414-60
; Sequence 60, Application US/09825414
; Patent No. US20020083489A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Collier, Alan
; APPLICANT: Altano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; FILE REFERENCE: 19603/3743
; CURRENT APPLICATION NUMBER: US/09/825,414
; PRIORITY FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 91
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Pseudomonas syringae pv. delphinii
US-09-825-414-60
```

```
Query Match          0.8%; Score 7; DB 10; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 819 IRGYAHD 825
|||||
DB 146 IRGYAHD 152
```

```
RESULT 22
US-09-866-050A-705
; Sequence 705, Application US/09866050A
; Publication No. US20030040471A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murrison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; PRIORITY FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 705
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Mouse
US-09-866-050A-705
```

```
Query Match          0.8%; Score 7; DB 9; Length 255;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 404 LQAVRAL 410
|||||
DB 76 LQAVRAL 82
```

```
RESULT 23
US-10-063-547-118
; Sequence 118, Application US/10063547
; Publication No. US20020182638A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Collin K.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C1
; CURRENT APPLICATION NUMBER: US/10/063,547
; PRIORITY FILING DATE: 2002-05-02
; Prior Application removed - see file wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 118
; LENGTH: 261
; TYPE: PRT
```

QY 166 KRLVARTL 172  
|||||  
Db 5 KRLVARTL 11

RESULT 16  
US-09-975-132A-27  
; Sequence 27, Application US/09975132A  
; Publication No. US20020182672A1  
; GENERAL INFORMATION:  
; APPLICANT: Kolman, Marc  
; TITLE OF INVENTION: Enhanced Secretion of a Polypeptide by a  
; FILE REFERENCE: GC636-2  
; CURRENT APPLICATION NUMBER: US/09/975,132A  
; PRIOR FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US 60/239,531  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 27  
; LENGTH: 158  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; OTHER INFORMATION: IL-3 substituted tag changed C-terminus  
US-09-975-132A-27

Query Match 0.8%; Score 7; DB 9; Length 158;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 KRLVARTL 172  
|||||  
Db 5 KRLVARTL 11

RESULT 17  
US-09-975-132A-28  
; Sequence 28, Application US/09975132A  
; Publication No. US20020182672A1  
; GENERAL INFORMATION:  
; APPLICANT: Kolman, Marc  
; TITLE OF INVENTION: Enhanced Secretion of a Polypeptide by a  
; FILE REFERENCE: GC636-2  
; CURRENT APPLICATION NUMBER: US/09/975,132A  
; PRIOR FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US 60/239,531  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 28  
; LENGTH: 173  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; OTHER INFORMATION: tagged IL-3 amino acid  
US-09-975-132A-28

Query Match 0.8%; Score 7; DB 9; Length 173;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 KRLVARTL 172  
|||||  
Db 5 KRLVARTL 11

RESULT 18  
US-09-986-480-368  
; Sequence 368, Application US/09986480  
; Publication No. US20030027999A1

; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 143 Human Secreted Proteins  
; FILE REFERENCE: PS500P1  
; CURRENT APPLICATION NUMBER: US/09/986,480  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: PCT/US00/12788  
; PRIOR FILING DATE: 2000-05-11  
; PRIOR APPLICATION NUMBER: US 60/134,068  
; PRIOR FILING DATE: 1999-05-13  
; NUMBER OF SEQ ID NOS: 456  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 368  
; LENGTH: 187  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-986-480-368

Query Match 0.8%; Score 7; DB 9; Length 187;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 384 PVKRELL 390  
|||||  
Db 31 PVKRELL 37

RESULT 19  
US-10-062-254-162  
; Sequence 162, Application US/10062254  
; Patent No. US20020138882A1  
; GENERAL INFORMATION:  
; APPLICANT: Cahoon, Edgar B  
; APPLICANT: Cahoon, Rebecca E  
; APPLICANT: Falco, Saverio Carl  
; APPLICANT: Fang, Yiwen  
; APPLICANT: Hantke, Sabine S.  
; APPLICANT: Lee, Jian-Ming  
; APPLICANT: Li, Zhongsen  
; APPLICANT: Miao, Guo-Hua  
; APPLICANT: Morgante, Michele  
; APPLICANT: Niu, Xiping  
; APPLICANT: Odell, Joan  
; APPLICANT: Rafalski, Antoni  
; APPLICANT: Sakai, Hajime  
; APPLICANT: Zheng, Peizhong  
; APPLICANT: Zhu, Qun  
; TITLE OF INVENTION: Polynucleotides Encoding Proteins Involved In Plant Metabolism  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/062,254  
; CURRENT FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: 09/630,346  
; PRIOR FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 60/146511  
; PRIOR FILING DATE: 1999-07-30  
; PRIOR APPLICATION NUMBER: 60/156006  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 60/156899  
; PRIOR FILING DATE: 1999-09-30  
; PRIOR APPLICATION NUMBER: 60/157287  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: 60/169767  
; PRIOR FILING DATE: 1999-12-09  
; PRIOR APPLICATION NUMBER: 60/171054  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: 60/172958  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 60/171515  
; PRIOR FILING DATE: 1999-12-22  
; PRIOR APPLICATION NUMBER: 60/173535  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 375  
; SOFTWARE: Microsoft Office 97

```

; CURRENT APPLICATION NUMBER: US/09/764,855
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 140
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-855-140

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 106;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 NKVPRLK 184
Db 69 NKVPRLK 75

RESULT 12
US-10-042-141-91
; Sequence 91, Application US/10042141
; Publication No. US20020183503A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/10/042,141
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: 09/726,643
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 91
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-141-91

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 122;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 DRVLAIn 514
Db 45 DRVLAIn 51

RESULT 13
US-09-726-643-91
; Sequence 91, Application US/09726643
; Patent No. US20020028449A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/09/726,643
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 91
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```

US-09-726-643-91

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 122;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 DRVLAIn 514
Db 45 DRVLAIn 51

RESULT 14
US-09-764-868-1129
; Sequence 1129, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P7232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1129
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (125)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (138)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-1129

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 150;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 DYNVFD 91
Db 85 DYNVFD 91

RESULT 15
US-09-975-132A-26
; Sequence 26, Application US/09975132A
; Publication No. US20020182672A1
; GENERAL INFORMATION:
; APPLICANT: Kolkman, Marc
; TITLE OF INVENTION: Enhanced Secretion of a Polypeptide by a
; FILE REFERENCE: GC636-2
; CURRENT APPLICATION NUMBER: US/09/975,132A
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US 60/239,531
; PRIOR FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: IL-3 encoded by plasmid pLATIL-3
US-09-975-132A-26

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 158;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```

; Sequence 24, Application US/09746919
; Patent No. US20020013452A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Howard M.
; APPLICANT: Pontzer, Carol H.
; TITLE OF INVENTION: Interferon Tau Compositions and
; TITLE OF INVENTION: Methods of Use
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Denlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/746,919
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/045,467
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/438,753
; FILING DATE: 10-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/139,891
; FILING DATE: 19-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/847,741
; FILING DATE: 09-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/318,050
; FILING DATE: 02-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/969,890
; FILING DATE: 30-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Denlinger, Peter J.
; REGISTRATION NUMBER: 28,006
; REFERENCE/DOCKET NUMBER: 5600-0001.36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE:
; US-09-746-919-24
; Query Match
; Best Local Similarity 0.8%; Score 7; DB 10; Length 95;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 389 LLEQLLT 395
| | | | |
Db 42 LLEQLLT 48

```

```

RESULT 9
; Sequence 506, Application US/09925297
; Patent No. US20020081659A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 506
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (45)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (92)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-506

```

```

Query Match
; Best Local Similarity 0.8%; Score 7; DB 10; Length 102;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 57 PVLITP 63
| | | | |
Db 65 PVLITP 71

```

```

RESULT 10
; Sequence 140, Application US/10072349
; Publication No. US20030054420A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA110C1
; CURRENT APPLICATION NUMBER: US/10/072,349
; PRIOR FILING DATE: 2002-02-11
; PRIOR APPLICATION removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: Patentin Ver. 3.1
; SEQ ID NO 140
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-349-140

```

```

Query Match
; Best Local Similarity 0.8%; Score 7; DB 9; Length 106;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 178 NKVPRLK 184
| | | | |
Db 69 NKVPRLK 75

```

```

RESULT 11
; Sequence 140, Application US/09764855
; Patent No. US20020119919A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA110

```

LENGTH: 41  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-726-643-93

Query Match 0.8%; Score 7; DB 10; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 DRYLAIN 514  
 |||||||  
 Db 4 DRYLAIN 10

RESULT 4  
 US-10-091-504-1204  
 ; Sequence 1204, Application US/10091504  
 ; Publication No. US20030059908A1  
 ; GENERAL INFORMATION:

APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PC007C1  
 CURRENT APPLICATION NUMBER: US/10/091,504  
 CURRENT FILING DATE: 2002-03-07  
 NUMBER OF SEQ ID NOS: 2442  
 Prior Application removed - See File Wrapper or Palm  
 SOFTWARE: Patentln Ver. 2.0  
 SEQ ID NO 1204  
 LENGTH: 51  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:

NAME/KEY: misc-feature  
 LOCATION: (20)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 NAME/KEY: misc-feature  
 LOCATION: (40)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 US-10-091-504-1204

Query Match 0.8%; Score 7; DB 9; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 LROTALV 240  
 |||||||  
 Db 25 LROTALV 31

RESULT 5  
 US-09-764-869-1204  
 ; Sequence 1204, Application US/09764869  
 ; Patent No. US20020061521A1  
 ; GENERAL INFORMATION:

APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PC007  
 CURRENT APPLICATION NUMBER: US/09/764,869  
 CURRENT FILING DATE: 2001-01-17  
 Prior Application data removed - refer to PALM or file wrapper  
 NUMBER OF SEQ ID NOS: 2442  
 SOFTWARE: Patentln Ver. 2.0  
 SEQ ID NO 1204  
 LENGTH: 51  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:

NAME/KEY: SITE  
 LOCATION: (20)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 NAME/KEY: SITE  
 LOCATION: (40)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-869-1204

Query Match 0.8%; Score 7; DB 10; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 LROTALV 240  
 |||||||  
 Db 25 LROTALV 31

RESULT 6

US-09-809-391-601  
 ; Sequence 601, Application US/09809391  
 ; Publication No. US20030049618A1  
 ; GENERAL INFORMATION:  
 APPLICANT: Ruben et al.  
 TITLE OF INVENTION: 186 Human Secreted proteins  
 FILE REFERENCE: P2002P2  
 CURRENT APPLICATION NUMBER: US/09/809,391  
 CURRENT FILING DATE: 2001-03-16  
 Prior application data removed - consult PALM or file wrapper  
 NUMBER OF SEQ ID NOS: 761  
 SOFTWARE: Patentln Ver. 2.0  
 SEQ ID NO 601  
 LENGTH: 69  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-809-391-601

Query Match 0.8%; Score 7; DB 9; Length 69;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 ELLEQL 394  
 |||||||  
 Db 51 ELLEQL 57

RESULT 7

US-09-809-391-465  
 ; Sequence 465, Application US/09809391  
 ; Publication No. US20030049618A1  
 ; GENERAL INFORMATION:  
 APPLICANT: Ruben et al.  
 TITLE OF INVENTION: 186 Human Secreted proteins  
 FILE REFERENCE: P2002P2  
 CURRENT APPLICATION NUMBER: US/09/809,391  
 CURRENT FILING DATE: 2001-03-16  
 Prior application data removed - consult PALM or file wrapper  
 NUMBER OF SEQ ID NOS: 761  
 SOFTWARE: Patentln Ver. 2.0  
 SEQ ID NO 465  
 LENGTH: 70  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:

NAME/KEY: SITE  
 LOCATION: (70)  
 OTHER INFORMATION: Xaa equals stop translation  
 US-09-809-391-465

Query Match 0.8%; Score 7; DB 9; Length 70;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 ELLEQL 394  
 |||||||  
 Db 51 ELLEQL 57

RESULT 8

US-09-746-919-24

```

969 6 0.7 101 9 US-09-738-626-6090 Sequence 6090, Ap
970 6 0.7 101 10 US-09-864-761-35856 Sequence 35556, A
971 6 0.7 104 9 US-09-834-794A-2 Sequence 2, Appl1
972 6 0.7 104 10 US-09-834-795A-2 Sequence 2, Appl1
973 6 0.7 105 9 US-10-023-934-89 Sequence 89, Appl1
974 6 0.7 106 9 US-10-023-896-82 Sequence 82, Appl1
975 6 0.7 107 10 US-09-864-761-43059 Sequence 43059, A
976 6 0.7 112 9 US-10-051-643-78 Sequence 78, Appl1
977 6 0.7 112 9 US-09-880-505-78 Sequence 78, Appl1
978 6 0.7 112 9 US-09-974-879-188 Sequence 188, Appl1
979 6 0.7 112 10 US-09-864-761-34124 Sequence 34124, A
980 6 0.7 113 10 US-09-925-300-1011 Sequence 1011, Ap
981 6 0.7 114 10 US-08-979-847-90 Sequence 90, Appl1
982 6 0.7 118 10 US-09-867-550-1670 Sequence 1670, Ap
983 6 0.7 119 9 US-10-101-464A-581 Sequence 581, Appl1
984 6 0.7 122 9 US-10-067-974-14 Sequence 14, Appl1
985 6 0.7 122 10 US-09-815-242-10421 Sequence 10421, A
986 6 0.7 122 10 US-09-815-242-13784 Sequence 13784, A
987 6 0.7 123 10 US-09-739-254-130 Sequence 130, App
988 6 0.7 123 10 US-09-904-615-130 Sequence 130, App
989 6 0.7 125 10 US-09-864-761-36218 Sequence 36218, A
990 6 0.7 125 9 US-09-738-626-6762 Sequence 6762, Ap
991 6 0.7 127 9 US-09-898-751A-6 Sequence 6, Appl1
992 6 0.7 127 9 US-09-738-626-5014 Sequence 5014, Ap
993 6 0.7 127 9 US-09-834-794A-1 Sequence 1, Appl1
994 6 0.7 127 9 US-10-146-496-2 Sequence 2, Appl1
995 6 0.7 127 10 US-09-813-492-2 Sequence 2, Appl1
996 6 0.7 127 10 US-09-834-795A-1 Sequence 1, Appl1
997 6 0.7 127 10 US-09-931-381A-2 Sequence 2, Appl1
998 6 0.7 130 9 US-09-925-299-1128 Sequence 1128, Ap
999 6 0.7 130 10 US-09-925-299-1128 Sequence 1128, Ap
1000 6 0.7 131 9 US-09-764-868-1156 Sequence 1156, Ap

```

## ALIGNMENTS

RESULT 1  
US-10-045-792-2

; Sequence 2, Application US/10045792  
; Publication No. US20030003563A1

GENERAL INFORMATION:

APPLICANT: Vinkemeier, Dwe  
Moarefi, Ismail  
Kurien, Jr., James E.

TITLE OF INVENTION: A CRYSTAL OF THE N-TERMINAL DOMAIN OF A  
Kuriyan, John

NUMBER OF SEQUENCES: 13  
STAT PROTEIN AND METHODS OF USE THEREOF

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION NUMBER: US/10/045,792

APPLICATION NUMBER: US/10/045,792

FILING DATE: 19-Oct-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/012,710

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-194

TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 851 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-045-792-2

```

```

Query Match 1.0%; Score 9; DB 9; Length 851;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 457 ESTLEPVIE 465
Db 756 ESTLEPVIE 764

```

RESULT 2

US-10-042-141-93

; Sequence 93, Application US/10042141

; Publication No. US20020183503A1

GENERAL INFORMATION:

APPLICANT: Ruben et al.

TITLE OF INVENTION: 26 Human secreted proteins

FILE REFERENCE: P2040P1

CURRENT APPLICATION NUMBER: US/10/042,141

PRIOR FILING DATE: 2002-01-11

PRIOR APPLICATION NUMBER: 09/726,643

PRIOR FILING DATE: 2000-12-01

PRIOR APPLICATION NUMBER: PCT/US00/15187

PRIOR FILING DATE: 2000-06-02

PRIOR APPLICATION NUMBER: 60/137,725

PRIOR FILING DATE: 1999-06-07

NUMBER OF SEQ ID NOS: 190

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 93

LENGTH: 41

TYPE: PRT

ORGANISM: Homo sapiens

US-10-042-141-93

Query Match 0.8%; Score 7; DB 9; Length 41;

Best Local Similarity 100.0%; Pred. No. 37;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 DRVLAIV 514

Db 4 DRVLAIV 10

RESULT 3

US-09-726-643-93

; Sequence 93, Application US/09726643

; Patent No. US20020028449A1

GENERAL INFORMATION:

APPLICANT: Ruben et al.

TITLE OF INVENTION: 26 Human secreted proteins

FILE REFERENCE: P2040P1

CURRENT APPLICATION NUMBER: US/09/726,643

PRIOR FILING DATE: 2000-12-01

PRIOR APPLICATION NUMBER: PCT/US00/15187

PRIOR FILING DATE: 2000-06-02

PRIOR APPLICATION NUMBER: 60/137,725

PRIOR FILING DATE: 1999-06-07

NUMBER OF SEQ ID NOS: 190

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 93

823	7	0.8	632	9	US-10-205-894-166	Sequence 166, App	896	7	0.8	984	9	US-10-101-464A-919	Sequence 919, App
824	7	0.8	632	9	US-10-205-896-166	Sequence 166, App	897	7	0.8	1014	9	US-10-101-464A-807	Sequence 807, App
825	7	0.8	632	9	US-10-205-898-166	Sequence 166, App	898	7	0.8	1021	9	US-10-101-464A-954	Sequence 954, App
826	7	0.8	632	9	US-10-205-901-166	Sequence 166, App	899	7	0.8	1226	9	US-09-117-447-2	Sequence 2, App1
827	7	0.8	632	9	US-10-205-903-166	Sequence 166, App	900	7	0.8	2336	12	US-10-033-026-10	Sequence 10, App1
828	7	0.8	632	9	US-10-206-809-166	Sequence 166, App	901	7	0.8	2969	9	US-09-738-626-4434	Sequence 4434, App
829	7	0.8	632	9	US-10-206-910-166	Sequence 166, App	902	6	0.7	9	9	US-09-264-516A-32	Sequence 32, App1
830	7	0.8	632	9	US-10-206-911-166	Sequence 166, App	903	6	0.7	9	9	US-09-909-460-95	Sequence 95, App1
831	7	0.8	632	9	US-10-206-912-166	Sequence 166, App	904	6	0.7	9	9	US-10-119-537-1	Sequence 1, App1
832	7	0.8	632	9	US-10-206-913-166	Sequence 166, App	905	6	0.7	10	10	US-09-894-018-327	Sequence 327, App
833	7	0.8	632	9	US-10-206-914-166	Sequence 166, App	906	6	0.7	13	10	US-09-791-378-430	Sequence 430, App
834	7	0.8	632	9	US-10-206-920-166	Sequence 166, App	907	6	0.7	15	9	US-09-865-989-252	Sequence 252, App
835	7	0.8	632	9	US-10-206-921-166	Sequence 166, App	908	6	0.7	15	9	US-10-099-574A-252	Sequence 252, App
836	7	0.8	632	9	US-10-206-923-166	Sequence 166, App	909	6	0.7	18	9	US-10-006-553-33	Sequence 33, App1
837	7	0.8	632	9	US-10-206-925-166	Sequence 166, App	910	6	0.7	19	9	US-09-834-794A-5	Sequence 5, App1
838	7	0.8	632	9	US-10-206-926-166	Sequence 166, App	911	6	0.7	19	10	US-09-834-795A-5	Sequence 5, App1
839	7	0.8	632	9	US-10-206-927-166	Sequence 166, App	912	6	0.7	22	9	US-09-933-999A-14	Sequence 14, App1
840	7	0.8	632	9	US-10-207-916-166	Sequence 166, App	913	6	0.7	24	9	US-10-001-883-134	Sequence 134, App
841	7	0.8	632	9	US-10-207-917-166	Sequence 166, App	914	6	0.7	24	10	US-09-864-761-12118	Sequence 42118, A
842	7	0.8	632	9	US-10-207-918-166	Sequence 166, App	915	6	0.7	25	9	US-10-097-065-284	Sequence 284, App
843	7	0.8	632	9	US-10-207-919-166	Sequence 166, App	916	6	0.7	27	10	US-09-864-761-35151	Sequence 35151, A
844	7	0.8	632	9	US-10-207-920-166	Sequence 166, App	917	6	0.7	32	9	US-10-174-410-209	Sequence 209, App
845	7	0.8	632	9	US-10-207-925-166	Sequence 166, App	918	6	0.7	32	10	US-09-821-883-11	Sequence 11, App1
846	7	0.8	632	9	US-10-208-021-166	Sequence 166, App	919	6	0.7	32	9	US-10-011-585A-159	Sequence 159, App
847	7	0.8	632	9	US-10-208-022-166	Sequence 166, App	920	6	0.7	37	9	US-09-864-761-39173	Sequence 39173, A
848	7	0.8	632	9	US-10-208-023-166	Sequence 166, App	921	6	0.7	38	10	US-09-925-302-747	Sequence 747, App
849	7	0.8	632	9	US-10-208-026-166	Sequence 166, App	922	6	0.7	40	10	US-09-864-761-36789	Sequence 36789, A
850	7	0.8	632	9	US-10-208-029-166	Sequence 166, App	923	6	0.7	40	10	US-09-986-480-212	Sequence 212, App
851	7	0.8	632	9	US-10-208-030-166	Sequence 166, App	924	6	0.7	41	9	US-09-864-761-33925	Sequence 33925, A
852	7	0.8	632	9	US-10-223-232-166	Sequence 166, App	925	6	0.7	46	10	US-09-864-761-46687	Sequence 46687, A
853	7	0.8	632	9	US-09-990-427-219	Sequence 219, App	926	6	0.7	48	10	US-09-864-761-43654	Sequence 43654, A
854	7	0.8	632	9	US-09-990-439-219	Sequence 219, App	927	6	0.7	50	10	US-09-864-761-44317	Sequence 44317, A
855	7	0.8	632	9	US-10-063-538-40	Sequence 40, App1	928	6	0.7	54	9	US-10-016-157A-176	Sequence 176, App
856	7	0.8	632	9	US-10-173-693-166	Sequence 166, App	929	6	0.7	55	9	US-10-091-572-293	Sequence 293, App
857	7	0.8	632	9	US-10-174-578-166	Sequence 166, App	930	6	0.7	59	9	US-10-119-537-2	Sequence 2, App1
858	7	0.8	632	9	US-10-175-741-166	Sequence 166, App	931	6	0.7	60	9	US-10-108-605-107	Sequence 107, App
859	7	0.8	632	9	US-10-175-750-166	Sequence 166, App	932	6	0.7	62	10	US-09-864-761-41663	Sequence 41663, A
860	7	0.8	632	9	US-10-176-986-166	Sequence 166, App	933	6	0.7	66	10	US-09-925-297-689	Sequence 689, App
861	7	0.8	632	9	US-10-184-664-166	Sequence 166, App	934	6	0.7	69	10	US-09-864-761-34228	Sequence 34228, A
862	7	0.8	632	9	US-10-187-888-166	Sequence 166, App	935	6	0.7	69	10	US-09-864-761-39101	Sequence 39101, A
863	7	0.8	632	9	US-10-194-360-166	Sequence 166, App	936	6	0.7	71	9	US-10-001-835-211	Sequence 211, App
864	7	0.8	632	9	US-10-194-365-166	Sequence 166, App	937	6	0.7	71	10	US-09-864-761-37218	Sequence 37218, A
865	7	0.8	632	9	US-10-195-895-166	Sequence 166, App	938	6	0.7	71	10	US-09-864-761-42991	Sequence 42991, A
866	7	0.8	632	9	US-10-195-898-166	Sequence 166, App	939	6	0.7	71	10	US-09-864-761-48254	Sequence 48254, A
867	7	0.8	632	9	US-10-196-759-166	Sequence 166, App	940	6	0.7	73	10	US-09-864-761-33893	Sequence 33893, A
868	7	0.8	632	9	US-10-199-302-166	Sequence 166, App	941	6	0.7	75	10	US-09-779-239-4	Sequence 4, App1
869	7	0.8	632	9	US-10-201-323-166	Sequence 166, App	942	6	0.7	77	10	US-09-864-761-40148	Sequence 40148, A
870	7	0.8	632	9	US-10-203-510-166	Sequence 166, App	943	6	0.7	77	10	US-09-867-550-574	Sequence 574, App
871	7	0.8	632	9	US-10-203-891-166	Sequence 166, App	944	6	0.7	77	10	US-09-796-336A-22	Sequence 587, App
872	7	0.8	632	9	US-10-205-904-166	Sequence 166, App	945	6	0.7	78	10	US-09-974-879-587	Sequence 37088, A
873	7	0.8	632	9	US-10-206-917-166	Sequence 166, App	946	6	0.7	79	9	US-09-335-325-11	Sequence 7, App1
874	7	0.8	632	9	US-10-207-923-166	Sequence 166, App	947	6	0.7	79	9	US-09-335-325-11	Sequence 11, App1
875	7	0.8	632	9	US-10-207-924-166	Sequence 166, App	948	6	0.7	79	9	US-10-131-241-7	Sequence 11, App1
876	7	0.8	632	9	US-10-208-028-166	Sequence 166, App	949	6	0.7	79	9	US-10-131-241-11	Sequence 11, App1
877	7	0.8	632	10	US-09-989-722-219	Sequence 219, App	950	6	0.7	79	10	US-09-761-120-7	Sequence 7, App1
878	7	0.8	632	10	US-09-989-723-219	Sequence 219, App	951	6	0.7	79	10	US-09-761-120-11	Sequence 11, App1
879	7	0.8	632	10	US-09-989-727-219	Sequence 219, App	952	6	0.7	80	10	US-09-864-761-48011	Sequence 48011, A
880	7	0.8	632	10	US-09-989-731-219	Sequence 219, App	953	6	0.7	80	10	US-09-925-297-554	Sequence 554, App
881	7	0.8	632	10	US-09-989-732-219	Sequence 219, App	954	6	0.7	81	10	US-10-080-960-20	Sequence 20, App1
882	7	0.8	632	10	US-09-991-073-219	Sequence 219, App	955	6	0.7	81	9	US-09-759-130B-60	Sequence 60, App1
883	7	0.8	632	10	US-09-991-073-219	Sequence 219, App	956	6	0.7	83	9	US-09-925-300-1336	Sequence 1336, App
884	7	0.8	632	10	US-09-990-442-219	Sequence 219, App	957	6	0.7	84	10	US-09-867-550-744	Sequence 744, App
885	7	0.8	632	10	US-09-991-163-219	Sequence 219, App	958	6	0.7	88	10	US-09-864-761-43431	Sequence 43431, A
886	7	0.8	632	10	US-09-993-604-219	Sequence 219, App	959	6	0.7	90	10	US-09-864-761-46509	Sequence 46509, A
887	7	0.8	632	10	US-09-990-456-219	Sequence 219, App	960	6	0.7	90	10	US-10-125-258-79	Sequence 79, App1
888	7	0.8	632	10	US-09-989-721-219	Sequence 219, App	961	6	0.7	92	9	US-10-125-258-80	Sequence 80, App1
889	7	0.8	632	12	US-10-006-867-40	Sequence 40, App1	962	6	0.7	92	9	US-09-877-843-27	Sequence 27, App1
890	7	0.8	632	12	US-10-052-586-166	Sequence 166, App	963	6	0.7	95	9	US-09-877-843-25	Sequence 25, App1
891	7	0.8	655	9	US-10-101-464A-70	Sequence 70, App1	964	6	0.7	97	9	US-09-877-843-26	Sequence 26, App1
892	7	0.8	656	9	US-09-764-868-959	Sequence 959, App	965	6	0.7	97	9	US-09-820-843A-55	Sequence 55, App1
893	7	0.8	749	10	US-09-899-569A-2	Sequence 2, App1	966	6	0.7	98	9	US-10-101-464A-654	Sequence 654, App
894	7	0.8	819	10	US-09-833-790-425	Sequence 425, App	967	6	0.7	98	9	US-10-083-357-827	Sequence 827, App
895	7	0.8	836	10	US-09-899-569A-4	Sequence 4, App1	968	6	0.7	99	9		

677	7	0.8	632	9	US-10-199-305-166	Sequence 166, App	750	7	0.8	632	9	US-10-205-893-166	Sequence 166, App
678	7	0.8	632	9	US-10-199-306-166	Sequence 166, App	751	7	0.8	632	9	US-10-205-897-166	Sequence 166, App
679	7	0.8	632	9	US-10-199-310-166	Sequence 166, App	752	7	0.8	632	9	US-09-993-469-219	Sequence 219, App
680	7	0.8	632	9	US-10-199-311-166	Sequence 166, App	753	7	0.8	632	9	US-09-993-748-219	Sequence 219, App
681	7	0.8	632	9	US-10-199-314-166	Sequence 166, App	754	7	0.8	632	9	US-09-997-442-219	Sequence 219, App
682	7	0.8	632	9	US-10-199-317-166	Sequence 166, App	755	7	0.8	632	9	US-10-053-567-40	Sequence 40, Appl
683	7	0.8	632	9	US-10-199-317-166	Sequence 166, App	756	7	0.8	632	9	US-10-174-571-166	Sequence 166, App
684	7	0.8	632	9	US-10-199-665-166	Sequence 166, App	757	7	0.8	632	9	US-10-176-746-166	Sequence 166, App
685	7	0.8	632	9	US-10-199-666-166	Sequence 166, App	758	7	0.8	632	9	US-10-176-746-166	Sequence 166, App
686	7	0.8	632	9	US-10-199-669-166	Sequence 166, App	759	7	0.8	632	9	US-10-183-011-166	Sequence 166, App
687	7	0.8	632	9	US-10-201-534-166	Sequence 166, App	760	7	0.8	632	9	US-10-184-639-166	Sequence 166, App
688	7	0.8	632	9	US-10-201-770-166	Sequence 166, App	761	7	0.8	632	9	US-10-187-742-166	Sequence 166, App
689	7	0.8	632	9	US-10-201-855-166	Sequence 166, App	762	7	0.8	632	9	US-10-187-748-166	Sequence 166, App
690	7	0.8	632	9	US-10-201-856-166	Sequence 166, App	763	7	0.8	632	9	US-10-187-748-166	Sequence 166, App
691	7	0.8	632	9	US-10-202-469-166	Sequence 166, App	764	7	0.8	632	9	US-10-188-771-166	Sequence 166, App
692	7	0.8	632	9	US-10-202-470-166	Sequence 166, App	765	7	0.8	632	9	US-10-188-771-166	Sequence 166, App
693	7	0.8	632	9	US-10-202-476-166	Sequence 166, App	766	7	0.8	632	9	US-10-192-008-166	Sequence 166, App
694	7	0.8	632	9	US-10-202-934-166	Sequence 166, App	767	7	0.8	632	9	US-10-192-009-166	Sequence 166, App
695	7	0.8	632	9	US-10-202-935-166	Sequence 166, App	768	7	0.8	632	9	US-10-192-012-166	Sequence 166, App
696	7	0.8	632	9	US-10-202-936-166	Sequence 166, App	769	7	0.8	632	9	US-10-192-014-166	Sequence 166, App
697	7	0.8	632	9	US-10-205-504-166	Sequence 166, App	770	7	0.8	632	9	US-10-192-016-166	Sequence 166, App
698	7	0.8	632	9	US-10-205-509-166	Sequence 166, App	771	7	0.8	632	9	US-10-192-016-166	Sequence 166, App
699	7	0.8	632	9	US-10-205-895-166	Sequence 166, App	772	7	0.8	632	9	US-10-194-362-166	Sequence 166, App
700	7	0.8	632	9	US-10-205-899-166	Sequence 166, App	773	7	0.8	632	9	US-10-194-364-166	Sequence 166, App
701	7	0.8	632	9	US-10-205-900-166	Sequence 166, App	774	7	0.8	632	9	US-10-194-395-166	Sequence 166, App
702	7	0.8	632	9	US-10-205-909-166	Sequence 166, App	775	7	0.8	632	9	US-10-194-424-166	Sequence 166, App
703	7	0.8	632	9	US-10-227-693-40	Sequence 40, Appl	776	7	0.8	632	9	US-10-194-458-166	Sequence 166, App
704	7	0.8	632	9	US-09-990-443-219	Sequence 219, App	777	7	0.8	632	9	US-10-194-459-166	Sequence 166, App
705	7	0.8	632	9	US-09-990-726-219	Sequence 219, App	778	7	0.8	632	9	US-10-194-488-	

531	7	0.8	632	9	US-10-180-555-166	Sequence 166, App	604	7	0.8	632	9	US-10-179-516-166	Sequence 166, App
532	7	0.8	632	9	US-10-180-559-166	Sequence 166, App	605	7	0.8	632	9	US-10-179-519-166	Sequence 166, App
533	7	0.8	632	9	US-10-181-000-166	Sequence 166, App	606	7	0.8	632	9	US-10-179-525-166	Sequence 166, App
534	7	0.8	632	9	US-10-183-010-166	Sequence 166, App	607	7	0.8	632	9	US-10-180-540-166	Sequence 166, App
535	7	0.8	632	9	US-10-183-012-166	Sequence 166, App	608	7	0.8	632	9	US-10-180-545-166	Sequence 166, App
536	7	0.8	632	9	US-10-184-614-166	Sequence 166, App	609	7	0.8	632	9	US-10-183-006-166	Sequence 166, App
537	7	0.8	632	9	US-10-184-623-166	Sequence 166, App	610	7	0.8	632	9	US-10-183-008-166	Sequence 166, App
538	7	0.8	632	9	US-10-184-635-166	Sequence 166, App	611	7	0.8	632	9	US-10-183-017-166	Sequence 166, App
539	7	0.8	632	9	US-10-184-637-166	Sequence 166, App	612	7	0.8	632	9	US-10-183-019-166	Sequence 166, App
540	7	0.8	632	9	US-10-184-646-166	Sequence 166, App	613	7	0.8	632	9	US-10-184-618-166	Sequence 166, App
541	7	0.8	632	9	US-10-184-647-166	Sequence 166, App	614	7	0.8	632	9	US-10-184-625-166	Sequence 166, App
542	7	0.8	632	9	US-10-184-652-166	Sequence 166, App	615	7	0.8	632	9	US-10-184-626-166	Sequence 166, App
543	7	0.8	632	9	US-10-187-594-166	Sequence 166, App	616	7	0.8	632	9	US-10-184-627-166	Sequence 166, App
544	7	0.8	632	9	US-10-187-596-166	Sequence 166, App	617	7	0.8	632	9	US-10-184-645-166	Sequence 166, App
545	7	0.8	632	9	US-10-187-745-166	Sequence 166, App	618	7	0.8	632	9	US-10-184-654-166	Sequence 166, App
546	7	0.8	632	9	US-10-187-885-166	Sequence 166, App	619	7	0.8	632	9	US-10-184-655-166	Sequence 166, App
547	7	0.8	632	9	US-10-187-886-166	Sequence 166, App	620	7	0.8	632	9	US-10-188-774-166	Sequence 166, App
548	7	0.8	632	9	US-10-199-464-166	Sequence 166, App	621	7	0.8	632	9	US-10-188-775-166	Sequence 166, App
549	7	0.8	632	9	US-10-176-751-166	Sequence 166, App	622	7	0.8	632	9	US-10-194-462-166	Sequence 166, App
550	7	0.8	632	9	US-10-176-760-166	Sequence 166, App	623	7	0.8	632	9	US-10-195-902-166	Sequence 166, App
551	7	0.8	632	9	US-10-176-990-166	Sequence 166, App	624	7	0.8	632	9	US-10-196-743-166	Sequence 166, App
552	7	0.8	632	9	US-10-180-541-166	Sequence 166, App	625	7	0.8	632	9	US-10-196-745-166	Sequence 166, App
553	7	0.8	632	9	US-10-180-542-166	Sequence 166, App	626	7	0.8	632	9	US-10-196-760-166	Sequence 166, App
554	7	0.8	632	9	US-10-180-548-166	Sequence 166, App	627	7	0.8	632	9	US-10-196-762-166	Sequence 166, App
555	7	0.8	632	9	US-10-180-551-166	Sequence 166, App	628	7	0.8	632	9	US-10-197-695-166	Sequence 166, App
556	7	0.8	632	9	US-10-180-998-166	Sequence 166, App	629	7	0.8	632	9	US-09-990-437-219	Sequence 219, App
557	7	0.8	632	9	US-10-180-999-166	Sequence 166, App	630	7	0.8	632	9	US-09-998-156-219	Sequence 166, App
558	7	0.8	632	9	US-10-183-013-166	Sequence 166, App	631	7	0.8	632	9	US-10-176-484-166	Sequence 166, App
559	7	0.8	632	9	US-10-184-612-166	Sequence 166, App	632	7	0.8	632	9	US-10-176-	

385	7	0.8	261	9	US-10-173-693-356	Sequence 356, App	458	7	0.8	516	10	US-09-883-797-14	Sequence 14, Appl
386	7	0.8	261	9	US-10-174-578-356	Sequence 356, App	459	7	0.8	534	9	US-09-975-719-67	Sequence 67, Appl
387	7	0.8	261	9	US-10-175-741-356	Sequence 356, App	460	7	0.8	599	9	US-09-738-626-5182	Sequence 5182, Ap
388	7	0.8	261	9	US-10-175-750-356	Sequence 356, App	461	7	0.8	632	9	US-09-928-598-219	Sequence 219, App
389	7	0.8	261	9	US-10-176-986-356	Sequence 356, App	462	7	0.8	632	9	US-09-989-293A-219	Sequence 219, App
390	7	0.8	261	9	US-10-184-641-356	Sequence 356, App	463	7	0.8	632	9	US-10-063-547-40	Sequence 40, Appl
391	7	0.8	261	9	US-10-187-888-356	Sequence 356, App	464	7	0.8	632	9	US-09-989-735-219	Sequence 219, App
392	7	0.8	261	9	US-10-194-360-356	Sequence 356, App	465	7	0.8	632	9	US-09-990-444-219	Sequence 219, App
393	7	0.8	261	9	US-10-194-365-356	Sequence 356, App	466	7	0.8	632	9	US-09-989-730-219	Sequence 219, App
394	7	0.8	261	9	US-10-195-895-356	Sequence 356, App	467	7	0.8	632	9	US-09-990-436-219	Sequence 219, App
395	7	0.8	261	9	US-10-195-898-356	Sequence 356, App	468	7	0.8	632	9	US-09-991-181-219	Sequence 219, App
396	7	0.8	261	9	US-10-196-759-356	Sequence 356, App	469	7	0.8	632	9	US-09-993-687-219	Sequence 219, App
397	7	0.8	261	9	US-10-199-302-356	Sequence 356, App	470	7	0.8	632	9	US-09-989-734-219	Sequence 219, App
398	7	0.8	261	9	US-10-201-323-356	Sequence 356, App	471	7	0.8	632	9	US-09-997-653-219	Sequence 219, App
399	7	0.8	261	9	US-10-205-510-356	Sequence 356, App	472	7	0.8	632	9	US-10-174-599-166	Sequence 166, App
400	7	0.8	261	9	US-10-203-891-356	Sequence 356, App	473	7	0.8	632	9	US-10-176-738-166	Sequence 166, App
401	7	0.8	261	9	US-10-203-904-356	Sequence 356, App	474	7	0.8	632	9	US-10-063-616-40	Sequence 40, Appl
402	7	0.8	261	9	US-10-206-917-356	Sequence 356, App	475	7	0.8	632	9	US-10-175-737-166	Sequence 166, App
403	7	0.8	261	9	US-10-207-923-356	Sequence 356, App	476	7	0.8	632	9	US-09-993-667-219	Sequence 219, App
404	7	0.8	261	9	US-10-207-924-356	Sequence 356, App	477	7	0.8	632	9	US-10-063-502-40	Sequence 40, Appl
405	7	0.8	261	12	US-10-208-028-356	Sequence 356, App	478	7	0.8	632	9	US-10-173-706-166	Sequence 166, App
406	7	0.8	261	12	US-10-006-867-118	Sequence 118, App	479	7	0.8	632	9	US-10-173-738-166	Sequence 166, App
407	7	0.8	261	12	US-10-052-586-356	Sequence 356, App	480	7	0.8	632	9	US-10-175-752-166	Sequence 166, App
408	7	0.8	297	9	US-09-981-353-90	Sequence 90, Appl	481	7	0.8	632	9	US-10-176-482-166	Sequence 166, App
409	7	0.8	308	9	US-09-738-626-6317	Sequence 509, App	482	7	0.8	632	9	US-10-176-757-166	Sequence 166, App
410	7	0.8	313	10	US-09-886-055-509	Sequence 10321, A	483	7	0.8	632	9	US-10-176-913-166	Sequence 166, App
411	7	0.8	323	10	US-09-815-242-10321	Sequence 10321, A	484	7	0.8	632	9	US-10-180-552-166	Sequence 166, App
412	7	0.8	323	10	US-09-815-242-13734	Sequence 13734, A	485	7	0.8	632	9	US-10-180-557-166	Sequence 166, App
413	7	0.8	343	9	US-10-028-072-162	Sequence 162, App	486	7	0.8	632	9	US-09-990-438-219	Sequence 219, App
414	7	0.8	343	9	US-10-121-049-162	Sequence 162, App	487	7	0.8	632	9	US-09-990-562-219	Sequence 219, App
415	7	0.8	343	9	US-10-123-904-162	Sequence 162, App	488	7	0.8	632	9	US-09-997-428-219	Sequence 219, App
416	7	0.8	343	9	US-10-140-470-162	Sequence 162, App	489	7	0.8	632	9	US-09-997-666-219	Sequence 219, App
417	7	0.8	343	9	US-10-175-746-162	Sequence 162, App	490	7	0.8	632	9	US-10-173-720-166	Sequence 166, App
418	7	0.8	343	9	US-10-176-918-162	Sequence 162, App	491	7	0.8	632	9	US-10-174-572-166	Sequence 166, App
419	7	0.8	343	9	US-10-137-865-162	Sequence 162, App	492	7	0.8	632	9	US-10-174-582-166	Sequence 166, App
420	7	0.8	343	9	US-10-140-928-162	Sequence 162, App	493	7	0.8	632	9	US-10-174-588-166	Sequence 166, App
421	7	0.8	343	9	US-10-142-431-162	Sequence 162, App	494	7	0.8	632	9	US-10-175-739-166	Sequence 166, App
422	7	0.8	343	9	US-10-143-114-162	Sequence 162, App	495	7	0.8	632	9	US-10-175-740-166	Sequence 166, App
423	7	0.8	343	9	US-10-142-419-162	Sequence 162, App	496	7	0.8	632	9	US-10-175-743-166	Sequence 166, App
424	7	0.8	343	9	US-10-142-419-162	Sequence 162, App	497	7	0.8	632	9	US-10-176-488-166	Sequence 166, App
425	7	0.8	343	9	US-10-142-419-162	Sequence 162, App	498	7	0.8	632	9	US-10-176-492-166	Sequence 166, App
426	7	0.8	343	9	US-10-142-419-162	Sequence 162, App	499	7	0.8	632	9	US-10-176-497-166	Sequence 166, App
427	7	0.8	343	9	US-10-142-419-162	Sequence 162, App	500	7	0.8	632	9	US-10-176-499-166	Sequence 166, App
428	7	0.8	343	9	US-10-142-419-162	Sequence 162, App	501	7	0.8	632	9	US-10-176-505-166	Sequence 166, App
429	7	0.8	343	9	US-10-141-755-162	Sequence 162, App	502	7	0.8	632	9	US-10-176-985-166	Sequence 166, App
430	7	0.8	343	9	US-10-143-032-162	Sequence 162, App	503	7	0.8	632	9	US-10-176-987-166	Sequence 166, App
431	7	0.8	343	9	US-10-123-108-162	Sequence 162, App	504	7	0.8	632	9	US-10-176-991-166	Sequence 166, App
432	7	0.8	343	9	US-10-123-236-162	Sequence 162, App	505	7	0.8	632	9	US-10-176-992-166	Sequence 166, App
433	7	0.8	343	9	US-10-140-928-162	Sequence 162, App	506	7	0.8	632	9	US-10-176-993-166	Sequence 166, App
434	7	0.8	343	9	US-10-121-045-162	Sequence 162, App	507	7	0.8	632	9	US-10-176-995-166	Sequence 166, App
435	7	0.8	343	9	US-10-123-292-162	Sequence 162, App	508	7	0.8	632	9	US-10-173-695-166	Sequence 166, App
436	7	0.8	343	9	US-10-123-292-162	Sequence 162, App	509	7	0.8	632	9	US-10-173-697-166	Sequence 166, App
437	7	0.8	343	9	US-10-123-292-162	Sequence 162, App	510	7	0.8	632	9	US-10-174-576-166	Sequence 166, App
438	7	0.8	343	9	US-10-124-819-162	Sequence 162, App	511	7	0.8	632	9	US-10-174-585-166	Sequence 166, App
439	7	0.8	343	9	US-10-124-822-162	Sequence 162, App	512	7	0.8	632	9	US-10-174-585-166	Sequence 166, App
440	7	0.8	343	9	US-10-140-925-162	Sequence 162, App	513	7	0.8	632	9	US-10-174-585-166	Sequence 166, App
441	7	0.8	343	9	US-10-160-498-162	Sequence 162, App	514	7	0.8	632	9	US-10-174-585-166	Sequence 166, App
442	7	0.8	343	10	US-09-815-242-11887	Sequence 11887, A	515	7	0.8	632	9	US-10-175-747-166	Sequence 166, App
443	7	0.8	349	10	US-09-916-051-1	Sequence 1, Appl1	516	7	0.8	632	9	US-10-176-481-166	Sequence 166, App
444	7	0.8	350	10	US-09-815-242-5712	Sequence 5712, Ap	517	7	0.8	632	9	US-10-176-487-166	Sequence 166, App
445	7	0.8	350	10	US-09-815-242-12472	Sequence 12472, A	518	7	0.8	632	9	US-10-176-493-166	Sequence 166, App
446	7	0.8	359	9	US-10-029-180-68	Sequence 68, Appl	519	7	0.8	632	9	US-10-176-493-166	Sequence 166, App
447	7	0.8	419	9	US-10-042-141-50	Sequence 50, Appl	520	7	0.8	632	9	US-10-176-911-166	Sequence 166, App
448	7	0.8	440	9	US-10-029-180-124	Sequence 124, Appl	521	7	0.8	632	9	US-10-176-919-166	Sequence 166, App
449	7	0.8	451	9	US-09-738-626-3548	Sequence 123, App	522	7	0.8	632	9	US-10-176-925-166	Sequence 166, App
450	7	0.8	452	9	US-09-815-242-5182	Sequence 3548, Ap	523	7	0.8	632	9	US-10-180-543-166	Sequence 166, App
451	7	0.8	453	9	US-09-939-825-21	Sequence 737, App	524	7	0.8	632	9	US-10-180-544-166	Sequence 166, App
452	7	0.8	491	10	US-09-738-626-6374	Sequence 6374, Ap	525	7	0.8	632	9	US-10-180-547-166	Sequence 166, App
453	7	0.8	494	9	US-10-199-922-2	Sequence 2, Appl1	526	7	0.8	632	9	US-10-180-547-166	Sequence 166, App
454	7	0.8	512	9			527	7	0.8	632	9	US-10-180-547-166	Sequence 166, App
455	7	0.8					528	7	0.8	632	9	US-10-180-547-166	Sequence 166, App
456	7	0.8					529	7	0.8	632	9	US-10-180-547-166	Sequence 166, App
457	7	0.8					530	7	0.8	632	9	US-10-180-549-166	Sequence 166, App





93	7	0.8	261	9	US-10-187-886-356	Sequence 356, App
94	7	0.8	261	9	US-10-199-464-356	Sequence 356, App
95	7	0.8	261	9	US-10-176-751-356	Sequence 356, App
96	7	0.8	261	9	US-10-176-760-356	Sequence 356, App
97	7	0.8	261	9	US-10-176-990-356	Sequence 356, App
98	7	0.8	261	9	US-10-180-541-356	Sequence 356, App
99	7	0.8	261	9	US-10-180-542-356	Sequence 356, App
100	7	0.8	261	9	US-10-180-543-356	Sequence 356, App
101	7	0.8	261	9	US-10-180-544-356	Sequence 356, App
102	7	0.8	261	9	US-10-180-551-356	Sequence 356, App
103	7	0.8	261	9	US-10-180-999-356	Sequence 356, App
104	7	0.8	261	9	US-10-183-013-356	Sequence 356, App
105	7	0.8	261	9	US-10-184-612-356	Sequence 356, App
106	7	0.8	261	9	US-10-184-616-356	Sequence 356, App
107	7	0.8	261	9	US-10-184-622-356	Sequence 356, App
108	7	0.8	261	9	US-10-184-623-356	Sequence 356, App
109	7	0.8	261	9	US-10-184-628-356	Sequence 356, App
110	7	0.8	261	9	US-10-184-630-356	Sequence 356, App
111	7	0.8	261	9	US-10-184-632-356	Sequence 356, App
112	7	0.8	261	9	US-10-184-633-356	Sequence 356, App
113	7	0.8	261	9	US-10-184-634-356	Sequence 356, App
114	7	0.8	261	9	US-10-184-635-356	Sequence 356, App
115	7	0.8	261	9	US-10-184-636-356	Sequence 356, App
116	7	0.8	261	9	US-10-184-650-356	Sequence 356, App
117	7	0.8	261	9	US-10-184-651-356	Sequence 356, App
118	7	0.8	261	9	US-10-187-588-356	Sequence 356, App
119	7	0.8	261	9	US-10-187-597-356	Sequence 356, App
120	7	0.8	261	9	US-10-187-598-356	Sequence 356, App
121	7	0.8	261	9	US-10-187-600-356	Sequence 356, App
122	7	0.8	261	9	US-10-187-601-356	Sequence 356, App
123	7	0.8	261	9	US-10-187-602-356	Sequence 356, App
124	7	0.8	261	9	US-10-187-603-356	Sequence 356, App
125	7	0.8	261	9	US-10-187-741-356	Sequence 356, App
126	7	0.8	261	9	US-10-187-743-356	Sequence 356, App
127	7	0.8	261	9	US-10-187-746-356	Sequence 356, App
128	7	0.8	261	9	US-10-187-747-356	Sequence 356, App
129	7	0.8	261	9	US-10-187-751-356	Sequence 356, App
130	7	0.8	261	9	US-10-187-753-356	Sequence 356, App
131	7	0.8	261	9	US-10-187-754-356	Sequence 356, App
132	7	0.8	261	9	US-10-187-757-356	Sequence 356, App
133	7	0.8	261	9	US-10-188-770-356	Sequence 356, App
134	7	0.8	261	9	US-10-188-773-356	Sequence 356, App
135	7	0.8	261	9	US-10-188-776-356	Sequence 356, App
136	7	0.8	261	9	US-10-188-779-356	Sequence 356, App
137	7	0.8	261	9	US-10-188-781-356	Sequence 356, App
138	7	0.8	261	9	US-10-188-783-356	Sequence 356, App
139	7	0.8	261	9	US-10-188-784-356	Sequence 356, App
140	7	0.8	261	9	US-10-188-785-356	Sequence 356, App
141	7	0.8	261	9	US-10-188-786-356	Sequence 356, App
142	7	0.8	261	9	US-10-188-787-356	Sequence 356, App
143	7	0.8	261	9	US-10-188-788-356	Sequence 356, App
144	7	0.8	261	9	US-10-188-789-356	Sequence 356, App
145	7	0.8	261	9	US-10-188-790-356	Sequence 356, App
146	7	0.8	261	9	US-10-188-791-356	Sequence 356, App
147	7	0.8	261	9	US-10-188-792-356	Sequence 356, App
148	7	0.8	261	9	US-10-188-793-356	Sequence 356, App
149	7					

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:35:40 ; Search time 48 Seconds  
(Without alignments)  
1534.152 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 919  
Sequence: 1 MSKRPVLEFANSFMPVALAAV.....TGVEEGNPILKHFICTPF 919

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 301932 seqs, 80129803 residues

Word size : 0

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

- 1: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubppaa/PC7\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubppaa/PC705\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.0	851	US-10-045-792-2	Sequence 2, Appl
2	7	0.8	41	US-10-042-141-93	Sequence 93, Appl
3	7	0.8	41	US-09-726-643-93	Sequence 93, Appl
4	4	0.8	51	US-10-091-504-1204	Sequence 1204, Ap
5	7	0.8	51	US-09-764-869-1204	Sequence 1204, Ap
6	7	0.8	69	US-09-809-391-601	Sequence 601, App
7	7	0.8	70	US-09-809-391-465	Sequence 465, App
8	7	0.8	95	US-09-746-919-24	Sequence 24, Appl
9	7	0.8	102	US-09-925-297-506	Sequence 506, App
10	7	0.8	106	US-10-072-349-140	Sequence 140, App
11	7	0.8	106	US-09-764-855-140	Sequence 140, App
12	7	0.8	122	US-10-042-141-91	Sequence 91, Appl
13	7	0.8	122	US-09-726-643-91	Sequence 91, Appl
14	7	0.8	150	US-09-764-868-1129	Sequence 1129, Ap
15	7	0.8	158	US-09-975-132A-26	Sequence 26, Appl
16	7	0.8	158	US-09-975-132A-27	Sequence 27, Appl
17	7	0.8	173	US-09-975-132A-28	Sequence 28, Appl
18	7	0.8	187	US-09-986-480-368	Sequence 368, App
19	7	0.8	188	US-10-062-254-162	Sequence 162, App

20	7	0.8	214	US-09-764-868-956	Sequence 956, App
21	7	0.8	215	US-09-825-414-60	Sequence 60, Appl
22	7	0.8	235	US-09-866-050A-705	Sequence 705, App
23	7	0.8	261	US-10-063-547-118	Sequence 118, App
24	7	0.8	261	US-10-174-590-356	Sequence 356, App
25	7	0.8	261	US-10-176-758-356	Sequence 356, App
26	7	0.8	261	US-10-063-616-118	Sequence 118, App
27	7	0.8	261	US-10-175-737-356	Sequence 356, App
28	7	0.8	261	US-10-063-502-118	Sequence 118, App
29	7	0.8	261	US-10-173-706-356	Sequence 356, App
30	7	0.8	261	US-10-175-738-356	Sequence 356, App
31	7	0.8	261	US-10-175-752-356	Sequence 356, App
32	7	0.8	261	US-10-176-483-356	Sequence 356, App
33	7	0.8	261	US-10-176-757-356	Sequence 356, App
34	7	0.8	261	US-10-176-913-356	Sequence 356, App
35	7	0.8	261	US-10-180-552-356	Sequence 356, App
36	7	0.8	261	US-10-180-557-356	Sequence 356, App
37	7	0.8	261	US-10-173-700-356	Sequence 356, App
38	7	0.8	261	US-10-174-572-356	Sequence 356, App
39	7	0.8	261	US-10-174-579-356	Sequence 356, App
40	7	0.8	261	US-10-174-582-356	Sequence 356, App
41	7	0.8	261	US-10-174-588-356	Sequence 356, App
42	7	0.8	261	US-10-175-739-356	Sequence 356, App
43	7	0.8	261	US-10-175-740-356	Sequence 356, App
44	7	0.8	261	US-10-176-488-356	Sequence 356, App
45	7	0.8	261	US-10-176-492-356	Sequence 356, App
46	7	0.8	261	US-10-176-747-356	Sequence 356, App
47	7	0.8	261	US-10-176-750-356	Sequence 356, App
48	7	0.8	261	US-10-176-985-356	Sequence 356, App
49	7	0.8	261	US-10-176-987-356	Sequence 356, App
50	7	0.8	261	US-10-176-991-356	Sequence 356, App
51	7	0.8	261	US-10-176-992-356	Sequence 356, App
52	7	0.8	261	US-10-176-993-356	Sequence 356, App
53	7	0.8	261	US-10-184-658-356	Sequence 356, App
54	7	0.8	261	US-10-173-695-356	Sequence 356, App
55	7	0.8	261	US-10-173-697-356	Sequence 356, App
56	7	0.8	261	US-10-173-705-356	Sequence 356, App
57	7	0.8	261	US-10-174-576-356	Sequence 356, App
58	7	0.8	261	US-10-174-585-356	Sequence 356, App
59	7	0.8	261	US-10-174-586-356	Sequence 356, App
60	7	0.8	261	US-10-175-747-356	Sequence 356, App
61	7	0.8	261	US-10-176-481-356	Sequence 356, App
62	7	0.8	261	US-10-176-485-356	Sequence 356, App
63	7	0.8	261	US-10-176-487-356	Sequence 356, App
64	7	0.8	261	US-10-176-493-356	Sequence 356, App
65	7	0.8	261	US-10-176-493-356	Sequence 356, App
66	7	0.8	261	US-10-176-756-356	Sequence 356, App
67	7	0.8	261	US-10-176-911-356	Sequence 356, App
68	7	0.8	261	US-10-176-919-356	Sequence 356, App
69	7	0.8	261	US-10-176-925-356	Sequence 356, App
70	7	0.8	261	US-10-176-978-356	Sequence 356, App
71	7	0.8	261	US-10-179-510-356	Sequence 356, App
72	7	0.8	261	US-10-180-543-356	Sequence 356, App
73	7	0.8	261	US-10-180-544-356	Sequence 356, App
74	7	0.8	261	US-10-180-546-356	Sequence 356, App
75	7	0.8	261	US-10-180-547-356	Sequence 356, App
76	7	0.8	261	US-10-180-549-356	Sequence 356, App
77	7	0.8	261	US-10-180-555-356	Sequence 356, App
78	7	0.8	261	US-10-180-559-356	Sequence 356, App
79	7	0.8	261	US-10-181-000-356	Sequence 356, App
80	7	0.8	261	US-10-183-010-356	Sequence 356, App
81	7	0.8	261	US-10-183-012-356	Sequence 356, App
82	7	0.8	261	US-10-184-614-356	Sequence 356, App
83	7	0.8	261	US-10-184-623-356	Sequence 356, App
84	7	0.8	261	US-10-184-635-356	Sequence 356, App
85	7	0.8	261	US-10-184-637-356	Sequence 356, App
86	7	0.8	261	US-10-184-646-356	Sequence 356, App
87	7	0.8	261	US-10-184-647-356	Sequence 356, App
88	7	0.8	261	US-10-184-652-356	Sequence 356, App
89	7	0.8	261	US-10-187-594-356	Sequence 356, App
90	7	0.8	261	US-10-187-596-356	Sequence 356, App
91	7	0.8	261	US-10-187-745-356	Sequence 356, App
92	7	0.8	261	US-10-187-885-356	Sequence 356, App

Tue Apr 29 12:22:27 2003

us-09-914-168-2.Oligo.rai

Page 24

Db 24 PLSUEEL 30

Search completed: April 28, 2003, 16:36:30  
Job time : 67 secs

```

; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 64
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Mouse
US-09-724-864-64

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 264;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 404 LQAVRAL 410
Db 76 LQAVRAL 82

RESULT 37
US-08-465-971B-2
; Sequence 2, Application US/08465971B
; Patent No. 5942414
; GENERAL INFORMATION:
; APPLICANT: YI LI and Mark D. Adams
; TITLE OF INVENTION: Human G-Protein Receptor HIBEF51
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: US
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,971B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLINS, J.G.
; REGISTRATION NUMBER: 33073
; REFERENCE/DOCKET NUMBER: 325800-453 (PFI187)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 349 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-971B-2

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 349;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 190 SSOSGET 196
Db 233 SSOSGET 239

RESULT 38
US-09-134-001C-5150
; Sequence 5150, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

```

```

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GFC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO: 5150
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5150

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 546 EVIDLPE 552
Db 321 EVIDLPE 327

RESULT 39
US-09-325-932A-191
; Sequence 191, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant dev
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 191
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-325-932A-191

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 355;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 752 GSSGLVS 758
Db 112 GSSGLVS 118

RESULT 40
5187089-2
; Patent No. 5187089
; APPLICANT: SCOTT, RANDY W.; GOLINI, FRED; MCGROGAN, MICHAEL
; TITLE OF INVENTION: PROTEASE NEXIN-I VARIANTS WHICH INHIBIT
; ELASTASE
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/542,484
; FILING DATE: 21-JUN-1990
; SEQ ID NO: 2
; LENGTH: 397
5187089-2

Query Match
Best Local Similarity 100.0%; Score 7; DB 6; Length 397;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 117 PLSTLEL 123

```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/245,511  
FILING DATE: 18-MAY-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/116,541  
FILING DATE: 01-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-069 CIP  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 175 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: Neisseria gonorrhoeae  
US-08-245-511-26

Query Match  
Best Local Similarity 0.8%; Score 7; DB 2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 354 DTGTQYR 360  
Db 11 DTGTQYR 17  
|||||||

RESULT 34  
US-08-600-993A-26  
Sequence 26, Application US/08600993A  
Patent No. 5981229  
GENERAL INFORMATION:  
APPLICANT: Masure, H Robert  
APPLICANT: Pearce, Barbara J  
APPLICANT: Tuomanen, Elaine  
TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND  
TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/600,993A  
FILING DATE: 1-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/245,511  
FILING DATE: 18-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/116,541  
FILING DATE: 01-SEP-1994

ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-069 US  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 175 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: Neisseria gonorrhoeae  
US-08-600-993A-26

Query Match  
Best Local Similarity 0.8%; Score 7; DB 2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 354 DTGTQYR 360  
Db 11 DTGTQYR 17  
|||||||

RESULT 35  
US-09-323-872A-12  
Sequence 12, Application US/09323872A  
Patent No. 6395539  
GENERAL INFORMATION:  
APPLICANT: Coschigano, Peter  
TITLE OF INVENTION: Compositions and Methods for Bioremediation  
FILE REFERENCE: OHU-03640  
CURRENT APPLICATION NUMBER: US/09/323,872A  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: 09/072,433  
PRIOR FILING DATE: 1998-05-04  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 12  
LENGTH: 211  
TYPE: PRT  
ORGANISM: Azorhizobium caulinodans  
US-09-323-872A-12

Query Match  
Best Local Similarity 0.8%; Score 7; DB 4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 721 ALLAGVA 727  
Db 197 ALLAGVA 203  
|||||||

RESULT 36  
US-09-724-864-64  
Sequence 64, Application US/09724864  
Patent No. 6380362  
GENERAL INFORMATION:  
APPLICANT: Watson, James D  
APPLICANT: Murison, James G.  
TITLE OF INVENTION: Polynucleotides, polypeptides expressed  
TITLE OF INVENTION: by the polynucleotides and methods for their use.  
FILE REFERENCE: 11000,105001  
CURRENT APPLICATION NUMBER: US/09/724,864  
CURRENT FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678  
PRIOR FILING DATE: 1999-12-23

```
Db      42 LLEOLLT 48

RESULT 31
US-09-045-467-24
; Sequence 24, Application US/09045467
; GENERAL INFORMATION:
; APPLICANT: Johnson, Howard M.
; Ponzet, Carol H.
; TITLE OF INVENTION: Interferon Tau Compositions and
; Methods of Use
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,467
; FILING DATE: 20-Mar-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/455,021
; FILING DATE: 31-MAY-1995
; APPLICATION NUMBER: US 08/438,753
; FILING DATE: 10-MAY-1995
; APPLICATION NUMBER: US 08/139,891
; FILING DATE: 19-OCT-1993
; APPLICATION NUMBER: US 07/847,741
; FILING DATE: 09-MAR-1992
; APPLICATION NUMBER: US 07/318,050
; FILING DATE: 02-MAR-1989
; APPLICATION NUMBER: US 07/969,890
; FILING DATE: 30-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Dehlinger, Peter J.
; REGISTRATION NUMBER: 28,006
; REFERENCE/DOCKET NUMBER: 5600-0001.36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: predicted amino acid coding sequence
; of SEQ ID NO:23 (HuIfntau7).
; SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-09-045-467-24

Query Match      0.8%: Score 7; DB 4; Length 95:
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      389 LLEOLLT 395
        |||||
Db      42 LLEOLLT 48

RESULT 32
US-08-796-792-2
; Sequence 2, Application US/08796792
; Patent No. 6087163

; GENERAL INFORMATION:
; APPLICANT: Gennaro, Maria L.
; APPLICANT: Lyashchenko, Konstantin P.
; APPLICANT: Manca, Claudia M.A.
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS
; TITLE OF INVENTION: SPECIFIC PROTEINS AND GENES, MIXTURES OF ANTIGENS AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 45 Rockefeller Plaza, Suite 2800
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/796,792
; FILING DATE: 06-FEB-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/011,364
; FILING DATE: 09-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hone, William J.
; REGISTRATION NUMBER: 26,739
; REFERENCE/DOCKET NUMBER: 07763/03301
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-765-5070
; TELEFAX: 212-258-2291
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 159 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-796-792-2

Query Match      0.8%: Score 7; DB 3; Length 159;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      14 PVALAAV 20
        |||||
Db      25 PVALAAV 31

RESULT 33
US-08-245-511-26
; Sequence 26, Application US/08245511
; Patent No. 5928900
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomenen, Elaine
; TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
; TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

Query Match 0.8%; Score 7; DB 2; Length 95;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 389 LLEQLLT 395  
|||||||  
Db 42 LLEQLLT 48

Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 389 LLEQLLT 395  
|||||||  
Db 42 LLEQLLT 48

RESULT 29  
US-08-455-524B-24  
; Sequence 24, Application US/08455524B  
; Patent No. 5942223  
; GENERAL INFORMATION:  
; APPLICANT: Bazer, Fuller W.  
; APPLICANT: Johnson, Howard M.  
; APPLICANT: Pontzer, Carol H.  
; APPLICANT: Olt, Troy L.  
; APPLICANT: Van Heeke, Gino  
; TITLE OF INVENTION: Interferon Tau Compositions and  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Ave., Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/455,524B  
; FILING DATE: 31-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/438,753  
; FILING DATE: 10-MAY-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/139,891  
; FILING DATE: 19-OCT-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/847,741  
; FILING DATE: 09-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/318,050  
; FILING DATE: 02-MAR-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/969,890  
; FILING DATE: 30-OCT-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 5600-0001.32  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-324-0880  
; TELEFAX: 415-324-0960  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 95 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: predicted amino acid coding sequence  
; INDIVIDUAL ISOLATE: of SEQ ID NO:23 (HUIFNTAU7).  
US-08-455-524B-24  
Query Match 0.8%; Score 7; DB 2; Length 95;

RESULT 30  
US-08-455-021B-24  
; Sequence 24, Application US/08455021B  
; GENERAL INFORMATION:  
; APPLICANT: Bazer, Fuller W.  
; APPLICANT: Johnson, Howard M.  
; APPLICANT: Pontzer, Carol H.  
; APPLICANT: Olt, Troy L.  
; APPLICANT: Van Heeke, Gino  
; APPLICANT: Imakawa, Kazuhito  
; TITLE OF INVENTION: Interferon Tau Compositions and  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Ave., Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/455,021B  
; FILING DATE: 31-MAY-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/139,891  
; FILING DATE: 19-OCT-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/847,741  
; FILING DATE: 09-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/318,050  
; FILING DATE: 02-MAR-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/969,890  
; FILING DATE: 30-OCT-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 5600-0001.31  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-324-0880  
; TELEFAX: 415-324-0960  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 95 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: predicted amino acid coding sequence  
; INDIVIDUAL ISOLATE: of SEQ ID NO:23 (HUIFNTAU7).  
US-08-455-021B-24  
Query Match 0.8%; Score 7; DB 2; Length 95;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 389 LLEQLLT 395  
|||||||

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/969,890  
FILING DATE: 30-OCT-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles R.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 5600-0001.30  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-324-0880  
TELEFAX: 415-324-0960  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 95 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: predicted amino acid coding sequence  
US-08-438-753B-24 of SEQ ID NO:23 (HuIFNtau7).

Query Match 0.8%; Score 7; DB 1; Length 95;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 389 LLEOLLT 395  
|||||  
DB 42 LLEOLLT 48

RESULT 27  
US-08-443-883A-24  
Sequence 24, Application US/08443883A  
Patent No. 5738845

GENERAL INFORMATION:  
APPLICANT: Bazer, Fuller W.  
APPLICANT: Johnson, Howard M.  
APPLICANT: Pontzer, Carol H.  
APPLICANT: Ott, Troy L.  
APPLICANT: Van Hecke, Gino  
APPLICANT: Imakawa, Kazuhito  
TITLE OF INVENTION: Interferon Tau Compositions and  
TITLE OF INVENTION: Methods of Use  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESS: Dehlinger & Associates  
STREET: 350 Cambridge Ave., Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/443,883A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/139,891  
FILING DATE: 19-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/847,741  
FILING DATE: 09-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/318,050  
FILING DATE: 02-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/969,890  
FILING DATE: 30-OCT-1992  
ATTORNEY/AGENT INFORMATION:

NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 5600-0001.30  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-324-0880  
TELEFAX: 415-324-0960  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 95 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: predicted amino acid coding sequence  
US-08-443-883A-24 of SEQ ID NO:23 (HuIFNtau7).

Query Match 0.8%; Score 7; DB 1; Length 95;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 389 LLEOLLT 395  
|||||  
DB 42 LLEOLLT 48

RESULT 28  
US-08-631-328-24  
Sequence 24, Application US/08631328  
Patent No. 5939266

GENERAL INFORMATION:  
APPLICANT: Johnson, Howard M.  
APPLICANT: Pontzer, Carol H.  
APPLICANT: Subramaniam, Prem S.  
TITLE OF INVENTION: Hybrid Interferon Compositions and  
TITLE OF INVENTION: Methods of Use  
NUMBER OF SEQUENCES: 55  
CORRESPONDENCE ADDRESS:  
ADDRESS: Dehlinger & Associates  
STREET: 350 Cambridge Ave., Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/631,328  
FILING DATE: 12-APR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/438,753  
FILING DATE: 10-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles R.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 5600-0001.34  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-324-0880  
TELEFAX: 415-324-0960  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 95 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: predicted amino acid coding sequence  
US-08-631-328-24 of SEQ ID NO:23 (HuIFNtau7).



```

CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: U18671; AAA98760.1; -
DR      EMBL: M97934; -; NOT_ANNOTATED_CDS.
DR      EMBL: S81491; AAB36226.1; -
DR      EMBL: S81491; AAB36227.1; ALT_SEQ.
DR      HSSP: P42224; 1BF5.
DR      TRANSFAC: T01494; -.
DR      Genew: HGNC:11363; STAT2.
DR      MIM: 600556; -.
DR      InterPro: IPR000980; SH2.
DR      InterPro: IPR001217; STAT.
DR      Pfam: PF00017; SH2; 1.
DR      Pfam: PF01017; STAT; 1.
DR      Pfam: PF02864; STAT_bind; 1.
DR      Pfam: PF02865; STAT_proc; 1.
DR      SMART: SM00252; SH2; 1.
DR      PROSITE: PS50001; SH2; 1.
DR      Transcription regulation; DNA-binding; Nuclear protein;
DR      Phosphorylation; SH2 domain; Alternative splicing.
KW      DOMAIN
FT      DOMAIN 572 667
FT      MOD_RES 690 690
FT      FT VARSPLIC 621 652
FT      VARSPLIC 621 652
FT      FT PHOSPHORYLATION (BY JAK) (BY SIMILARITY).
FT      FT DKLIVSYQPYTKELQSLPTEIIRHQLTF -> GOLLC
FT      FT PAIPVPSPECLAPLLPCLPLPLASLALN (IN SHORT
FT      FT ISOFORM)
FT      FT MISSING (IN SHORT ISOFORM).
FT      FT E4C74674CB7A3215 CRC64;
SQ      SEQUENCE 851 AA; 97916 MW; E4C74674CB7A3215 CRC64;

Query Match 1.0%; Score 9; DB 1; Length 851;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 457 ESTLEPVTE 465
Db 756 ESTLEPVTE 764

RESULT 3
NARC_ECOLI
ID NARC_ECOLI STANDARD; PRT; 200 AA.
AC P33932;
DT 01-FEB-1994 (Rel. 28, Created)
DI 01-FEB-1994 (Rel. 28, Last sequence update)
DE 16-OCT-2001 (Rel. 40, last annotation update)
DE Cytochrome c-type protein napc.
GN NAPC OR B2202 OR Z3459 OR ECS3091.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.,
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
```

RX	MEDLINE-21074935; PubMed-11206551;
RA	Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA	Postfil G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA	Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouls K.,
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA	Weich R.A., Blatter F.R.;
RT	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL	Nature 409:529-533(2001).
RN	[4]
RP	SEQUENCE FROM N.A.
RC	STRAIN-O157:H7 / RIMD 0509952;
RX	MEDLINE-21156231; PubMed-11258796;
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA	Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA	Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA	Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT	*Complete genome sequence of enterohaemorrhagic Escherichia coli
RT	O157:H7 and genomic comparison with a laboratory strain K-12.";
RL	DNA Res. 8:11-22(2001).
RN	[5]
RP	CHARACTERIZATION AS A CYTOCHROME C.
RX	MEDLINE-94314186; PubMed-8039676;
RA	Tobbi-Niyol C., Crooke H., Griffiths L., Grov J., Hussain H.,
RT	Fommier J., Mejean V., Cole J.A.;
RT	"A reassessment of the range of c-type cytochromes synthesized by
RT	Escherichia coli K-12.";
RL	FEMS Microbiol. Lett. 119:89-94(1994).
CC	- FUNCTION: MEDIATES ELECTRON FLOW FROM QUINONES TO THE NAPAB
CC	COMPLEX.
CC	- SUBCELLULAR LOCATION: MEMBRANE-ANCHORED, PERIPLASMIC (POTENTIAL).
CC	- PTM: BINDS FOUR HEME GROUPS PER MOLECULE (POTENTIAL).
CC	- SIMILARITY: TO OTHER NAPC AND TO P.SHUTTER1 NRIT.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on tests
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC	or send an email to license@isb-sdb.ch).
CC	-----
DR	EMBL; U00008; AAA16394.1; -;
DR	EMBL; AE000309; AAC75262.1; -;
DR	EMBL; AE005452; AAG57337.1; -;
DR	EMBL; AP002560; BAB36514.1; -;
DR	EcoGene; BG12060; napc.
DR	InterPro; IPR000345; Cytc_heme_bind.
DR	InterPro; IPR005126; Cyt_NNT.
DR	Pfam; PF03264; Cytochrome_NNT; 1.
DR	PROSITE; PS00190; CYTOCHROME_C; 3.
KW	Electron transport; Heme; Transmembrane; Periplasmic;
KW	Complete proteome.
FT	DOMAIN 1 23
FT	TRANSMEM 24 44
FT	DOMAIN 45 200
FT	BINDING 57 57
FT	BINDING 60 60
FT	BINDING 61 61
FT	METAL 61 61
FT	FT
FT	BINDING 87 87
FT	BINDING 90 90
FT	FT
FT	METAL 91 91
FT	FT
FT	BINDING 147 147
FT	BINDING 150 150
FT	METAL 151 151
FT	FT
FT	BINDING 179 179
FT	BINDING 182 182
FT	METAL 183 183
FT	FT
SEQUENCE	200 AA; 23100 MW; EB6A62ABE46AA3 CRC64;
SEQUENCE	(BY SIMILARITY).

```

Query Match          0.9%; Score 8; DB 1; Length 200;
Best Local Similarity 100.0%; Pred.No.3.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 161 KPGLIKRL 168
    |||||||
Db 7 KPGLIKRL 14

RESULT 4
KAG2_CAEEL          STANDARD; PRT; 360 AA.
ID KAG2_CAEEL
AC 027535;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Probable arginine kinase ZC434.8 (EC 2.7.3.3) (AK).
GN ZC434.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Wilkinson J.;
RL Submitted (Jul-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ATP:GUANIDO PHOSPHOTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z75714; CAB0062.1; -
DR HSSP: P51541; IBC0
DR WormPep: ZC434.8; CE06583.
DR InterPro: IPR000749; ATP-gua_Ptrans.
DR Pfam: PF00217; ATP-gua_Ptrans. 1.
DR Pfam: PF02807; ATP-gua_Ptrans. 1.
DR ProSITE: PS00112; GUANIDO_KINASE. 1.
KW Hypothetical protein; Transferase; Kinase.
FT ACT_SITE 275 275
FT SEQUENCE 360 AA; 40382 MW; 176760B381EF26D7 CRC64;

Query Match          0.9%; Score 8; DB 1; Length 360;
Best Local Similarity 100.0%; Pred.No.6.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 492 LNLVAKA 499
    |||||||
Db 254 LNLVAKA 261

RESULT 5
YTFM_ECOLI          STANDARD; PRT; 577 AA.
ID YTFM_ECOLI
AC P39320;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein YTFM precursor.
GN YTFM OR B4220 OR Z5831 OR ECS5198.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562, 83334;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfal G., Hackett J., Klink S., Boulton A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [4]
RP -1- SIMILARITY: STRONG, TO H.INFLUENZAE H10698.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U14003; AAA97116.1; -
DR EMBL: AE000493; AAC77177.1; -
DR EMBL: AE005654; AAC59418.1; -
DR EMBL: AP002568; BAB38621.1; -
DR EcoGene: EG12513; YTFM.
DR InterPro: IPR000184; Bac_surfAq_D15.
DR Pfam: PF01103; Bac_surfAq_Ag. 1.
KW Hypothetical protein; Signal; complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 577
FT HYPOTHETICAL PROTEIN YTFM.
FT SEQUENCE 577 AA; 64796 MW; 253D1D5BE744D25 CRC64;

Query Match          0.9%; Score 8; DB 1; Length 577;
Best Local Similarity 100.0%; Pred.No.10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 809 LREFAGD 816
    |||||||
Db 466 LREFAGD 473

RESULT 6
MCCA_ARATH          STANDARD; PRT; 734 AA.
ID MCCA_ARATH
AC 042523; O9SA61;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Methylcrotonyl-CoA carboxylase alpha chain, mitochondrial precursor
DE (EC 6.4.1.4) (3-Methylcrotonyl-CoA carboxylase 1) (MCase alpha
DE subunit) (3-methylcrotonyl-CoA:carbon dioxide ligase alpha subunit).

```

GN MCCA OR AT1G03090 OR F1003\_8 OR F1003\_9.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=95232183; PubMed=7716229;  
RA Weaver L.M., Lebrun L., Franklin A., Huang L., Hoffman N.,  
RT Murrele E.S., Nikolau B.J.;  
RT "Molecular cloning of the biotinylated subunit of 3-methylcrotonyl-  
coenzyme A carboxylase of Arabidopsis thaliana.";  
RL Plant Physiol. 107:1013-1014(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=21016719; PubMed=11130712;  
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,  
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
RA Kim C.J., Koo H.U., Kremetskaia I., Kutz D.B., Kwan A., Lam B.,  
RA Landin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Maritali A.,  
RA Miltscher J., Miranda M., Nguyen N., Nierman W.C., Osborne B.I.,  
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
RA Wu D., Xu G., Fraser C.M., Venier J.C., Davis R.W.;  
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
thaliana.";  
RL Nature 408:816-820(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;  
RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the  
RT SSP consortium (Salk/Stanford/PGEC).";  
RL submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.  
RN [4]  
RP TISSUE SPECIFICITY.  
RC STRAIN=cv. Landsberg erecta, and cv. Columbia;  
RX MEDLINE=20148760; PubMed=10681539;  
RA McKeon A.L., Ke J., Song J., Che P., Achenbach S., Nikolau B.J.,  
RA Murrele E.S.;  
RT "Molecular characterization of the non-biotin-containing subunit of  
RT 3-methylcrotonyl-CoA carboxylase.";  
RL J. Biol. Chem. 275:5582-5590(2000).  
CC -1- CATALYTIC ACTIVITY: ATP + 3-methylcrotonyl-CoA + HCO(3)(-) = ADP +  
CC phosphate + 3-methylglutaconyl-CoA.  
CC -1- COFACTOR: Biotin.  
CC -1- PATHWAY: leucine catabolism.  
CC -1- SUBUNIT: Probably a heterodimer composed of biotin-containing  
CC alpha subunits and beta subunits (By similarity).  
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.  
CC -1- TISSUE SPECIFICITY: In roots, cotyledons, leaves, flowers,  
CC ovaries, siliques and embryos.  
CC -1- MISCELLANEOUS: Temporal and spatial accumulation of the alpha and  
CC beta subunits during development at approximately equal molar  
CC ratios.  
CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous  
CC gene model prediction.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U12536; AAA67356.1; -.
CC DR EMBL; AC006550; AAD25800.1; ALT_SEQ.
CC DR EMBL; AY070723; AAL50065.1; -.
CC DR HSSP; P24182; 1BNC.
CC DR InterPro; IPR001882; Biotin_attach.
CC DR InterPro; IPR000089; Biotin_11poy1.
CC DR InterPro; IPR000901; CPSase.
CC DR Pfam; PF02785; Biotin-carb_C; 1.
CC DR Pfam; PF00364; Biotin_11poy1; 1.
CC DR Pfam; PF00289; CPSase_L_chain; 2.
CC DR Pfam; PF02786; CPSase_L_D2; 2.
CC DR PROSITE; PS00188; BIOTIN; 1.
CC DR PROSITE; PS00866; CPSASE_1; 1.
CC DR PROSITE; PS00867; CPSASE_2; 1.
CC KW Mitochondrion; Ligase; Biotin; ATP-binding; Translt peptide.
CC FT TRANSIT 1 25 MITOCHONDRION (POTENTIAL).
CC FT CHAIN 26 734 METHYLCROTONYL-COA CARBOXYLASE ALPHA
CC FT NP_BIND 198 203 ATP (POTENTIAL).
CC FT ACT_SITE 329 329 BY SIMILARITY.
CC FT BINDING 699 699 BIOTIN (BY SIMILARITY).
CC FT CONFLICT 85 85 V -> D (IN REF. 1).
CC FT CONFLICT 92 92 A -> AK (IN REF. 1).
CC FT CONFLICT 281 300 MISSING (IN REF. 1).
CC FT CONFLICT 430 430 W -> L (IN REF. 1).
CC SQ SEQUENCE 734 AA; 80451 MW; 251CACF6464B046B CRC64;

Query Match 0.9%; Score 8; DB 1; Length 734;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 AARAVGY 248
| | | | | | | |
Db 298 AARAVGY 305

RESULT 7
CARR_THEAC
ID CARR_THEAC STANDARD; PRT; 1047 AA.
AC 09HH17:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate phosphatase ammonia chain).
DE GN CARR OR TAO791.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OC NCB1_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Gimpl W., Santos-Matinez M.-L., Koretke K.K., Volker C.,
RA Meves H.-W., Frisshman D., Stocker S., Lupas A.N., Baumeister W.;
RT The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum "J".
RL Nature 407:508-513(2000).
RC -1 CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1 COFACTOR: Binds three manganese ions (By similarity).
CC -1 PATHWAY: Arginine biosynthesis.
CC -1 SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -1 SIMILARITY: BELONGS TO THE CARR FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL445065; CAC11922.1; -
CC HSSP: P00968; 1JDB.
CC InterPro: IPR005483; CPase_L.
CC InterPro: IPR005479; CPase_L_D2.
CC InterPro: IPR005480; CPase_L_D3.
CC InterPro: IPR005481; CPase_L_N.
CC InterPro: IPR004362; MGS-like.
CC Pfam: PF00289; CPase_L_D2; 1.
CC Pfam: PF02786; CPase_L_D3; 1.
CC Pfam: PF02787; CPase_L_D3; 1.
CC Pfam: PF02142; MGS; 1.
CC PRINTS: PR00098; CPASASE.
CC PROSITE: PS00866; CPASE_1; FALSE_NEG.
CC PROSITE: PS00867; CPASE_2; 2.
CC Arginine biosynthesis: Pyrimidine biosynthesis: Ligase: Repeat;
CC ATP-binding: Manganese: Complete proteome.
CC CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
CC OLIGOMERIZATION DOMAIN.
CC CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
CC ALLOSTERIC DOMAIN.
CC DOMAIN 1 398
CC DOMAIN 399 539
CC DOMAIN 540 915
CC DOMAIN 916 1047
CC REPEAT 1 539
CC REPEAT 540 1047
CC NP_BIND 153 210
CC NP_BIND 300 350
CC METAL 284 284
CC METAL 296 296
CC METAL 298 298
CC METAL 813 813
CC METAL 825 825
CC SEQUENCE 1047 AA; 116851 MW; E062CAD131746864 CRC64;

Query Match 0.9%; Score 8; DB 1; Length 1047;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 626 KGVKLYAT 633
DB 941 KGVKLYAT 948

RESULT 8
SALM_DROME STANDARD: PRT: 1355 AA.
AC P39770;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeotic protein spalt-major.
GN SALM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94139659; PubMed=7905822;
RA Kuehnlein R.P., Frommer G., Friedrich M., Gonzalez-Gaitan M.,
RA Weber A., Wagner-Bernholz J.F., Gehring W.J., Jaechle H., Schuh R.;
RT "spalt encodes an evolutionarily conserved zinc finger protein of
RT novel structure which provides homeotic gene function in the head and
RT tail region of the Drosophila embryo.";
RL EMBL J. 13:168-179(1994).
CC -!- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF THE POSTERIOR-MOST

```

```

CC HEAD AND THE ANTERIOR-MOST TAIL SEGMENTS OF THE EMBRYO. PROBABLY
CC FUNCTION AS A TRANSCRIPTIONAL REGULATOR. COULD REPRESS THE
CC TRANSCRIPTION OF THE TSH GENE.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable)
CC -!- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT BLASTODERM STAGE AND
CC LATER IN RESTRICTED AREAS OF THE EMBRYONIC NERVOUS SYSTEM AS
CC WELL AS IN THE DEVELOPING TRACHEA.
CC -!- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X7541; CAA53229.1; -
CC HSSP: P15822; 1BBO.
CC Flybase: FBgn004579; salm.
CC InterPro: IPR000822; Znf.C2H2.
CC Pfam: PF00096; zfc-C2H2; 7.
CC PRINTS: PR00048; ZINC-FINGER.
CC SMART: SM00355; Znf.C2H2; 7.
CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
CC PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.
CC Developmental protein: Zinc-finger: Metal-binding: DNA-binding;
CC Nuclear protein: Transcription regulation; Repeat.
CC ZN_FING 451 473
CC ZN_FING 479 501
CC ZN_FING 824 846
CC ZN_FING 852 874
CC ZN_FING 884 906
CC ZN_FING 1289 1311
CC ZN_FING 1317 1339
CC SEQUENCE 1355 AA; 148995 MW; 6FC4EBDC0BC6355E CRC64;

Query Match 0.9%; Score 8; DB 1; Length 1355;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 LTPEDQIA 68
DB 913 LTPEDQIA 920

RESULT 9
SALM_DROVI STANDARD: PRT: 1402 AA.
AC P39606;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeotic protein spalt-major.
GN SALM.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7244;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94139659; PubMed=7905822;
RA Kuehnlein R.P., Frommer G., Friedrich M., Gonzalez-Gaitan M.,
RA Weber A., Wagner-Bernholz J.F., Gehring W.J., Jaechle H., Schuh R.;
RT "spalt encodes an evolutionarily conserved zinc finger protein of
RT novel structure which provides homeotic gene function in the head and
RT tail region of the Drosophila embryo.";
RL EMBL J. 13:168-179(1994).
CC -!- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF THE POSTERIOR-MOST
CC HEAD AND THE ANTERIOR-MOST TAIL SEGMENTS OF THE EMBRYO. PROBABLY
CC FUNCTION AS A TRANSCRIPTIONAL REGULATOR. COULD REPRESS THE

```

```

CC TRANSCRIPTION OF THE TSH GENE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -2- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT BLASTODERM STAGE AND
CC LATER IN RESTRICTED AREAS OF THE EMBRYONIC NERVOUS SYSTEM AS
CC WELL AS IN THE DEVELOPING TRACHEA.
CC -3- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z27444; CAA81800.1; -.
DR PIR: S42748; S42748.
DR HSSP: P15822; 1BBO.
DR FLYBASE: FBgn001317; Dvir\salin.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2.7.
DR PRINTS: PR00048; ZINC_FINGER.
DR SMART: SM00355; Znf_C2H2.7.
DR PROSITE: PS00026; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 7.
KW Developmental protein; zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Transcription regulation; Repeat.
FT ZN_FING 432 454 C2H2-TYPE.
FT ZN_FING 460 482 C2H2-TYPE.
FT ZN_FING 845 867 C2H2-TYPE.
FT ZN_FING 873 895 C2H2-TYPE.
FT ZN_FING 905 927 C2H2-TYPE.
FT ZN_FING 1336 1358 C2H2-TYPE.
FT ZN_FING 1364 1386 C2H2-TYPE.
SO SEQUENCE 1402 AA; 153745 MW; 4CDB2E384F70A5 CRC64;

Query Match 0.98; Score 8; DB 1; Length 1402;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 LTPRO10A 68
Db 934 LTPRO10A 941

RESULT 10
HYBG_ECOLI STANDARD; PRT; 82 AA.
ID HYBG_ECOLI
AC P37185;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hydrogenase-2 operon protein hydg.
GN HYBG OR B2990 OR Z4344 OR ECS3875.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / TG1;
RX MEDLINE-94292472; PubMed-8021226;
RA Menon N.K., Chatelus C.Y., Deravartanian M., Wendt J.C.,
RA Shammugam K.T., Beck H.D. Jr., Przytyla A.E.;
RT Cloning, sequencing, and mutational analysis of the hyd operon
RT encoding Escherichia coli hydrogenase 2."
RL J. Bacteriol. 176:4416-4423(1994).
[2]
RN RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97426617; PubMed-9278503;

```

```

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
[3]
RN RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE-21074935; PubMed-11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamotis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
[4]
RN RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE-21156231; PubMed-11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhnara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -1- FUNCTION: MAY HAVE A SPECIFIC ROLE IN THE MATURATION OF THE LARGE
CC SUBUNITS OF HYD1 AND HYD2.
CC -2- SIMILARITY: BELONGS TO THE HUPE/HYPC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U09177; AAA21595.1; -.
DR EMBL: U28377; AAA69157.1; -.
DR EMBL: AE000382; AAC76026.1; -.
DR EMBL: AE005529; AAG58127.1; -.
DR EMBL: AP002563; BAB37298.1; -.
DR Ecogen: EG11805; hydg.
DR InterPro: IPR001109; HUPE_HYPC.
DR Pfam: PF01455; HUPE_HYPC.1.
DR PRINTS: PR00445; HUPEHYPC.
DR PRODOM: PD003112; HUPE_HYPC.1.
DR TIGRFAMs: TIGR00074; hupC_hupF.1.
DR PROSITE: PS01097; HUPE_HYPC.1.
KW Complete proteome.
SQ SEQUENCE 82 AA; 8808 MW; 8E9736BFC4EFD626 CRC64;

Query Match 0.88; Score 7; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 839 GOVLAVG 845
Db 7 GOVLAVG 13

RESULT 11
YGT2_YEAST STANDARD; PRT; 107 AA.
ID YGT2_YEAST
AC P51103;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 12.2 kDa protein in COX4-GTSL intergenic region.

```

```

GN YG1182C OR G1607.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / FY1679;
RX MEDLINE=97197971; PubMed=9046087;
RA Coglievina M., Klima R., Bertani I., Delneri D., Zaccaria P.,
RT Brusch C.V.;
RT "Sequencing of a 40.5 kb fragment located on the left arm of
RT chromosome VII from Saccharomyces cerevisiae.";
RL Yeast 13:55-64(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X91489; CAA62792.1;
DR EMBL: 272704; CAA96894.1;
DR SGD: S0003150; YG1182C.
KW Hypothetical protein.
SO SEQUENCE 107 AA; 12220 MW; 1D22E64BC819A499 CRC64;

Query Match          0.8%; Score 7; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 687 YSLRYRL 693
   |||||
DB 84 YSLRYRL 90

RESULT 12
Y353_MYCPN          STANDARD; PRT; 109 AA.
AC P75249;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG353 homolog (G12_orf109).
GN MPM529 OR MP13.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkle E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000028; AAB95961.1;
DR InterPro: IPR000119; Bac_DNABind.
DR SMART: SM00411; BHL; 1.
KW Hypothetical protein; Complete proteome.
SO SEQUENCE 109 AA; 12410 MW; 3FF862745F3C5316 CRC64;

```

```

Query Match          0.8%; Score 7; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 898 VATGVKE 904
   |||||
DB 22 VATGVKE 28

RESULT 13
MP63_MYCTU          STANDARD; PRT; 159 AA.
AC P97175; 008224;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Immunogenic protein MP63/MPB63 precursor (Antigen MP63/MPB63) (16
DE kDa immunoprotective extracellular protein).
GN MP63 OR MPB63 OR RV1926C OR MT1977 OR MTCY09F9.38.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinobacteriales;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=97130011; PubMed=8975887;
RA Mance C.M.A., Lyashchenko K., Wiker H.G., Usai D., Colanigeli R.,
RA Genaro M.L.;
RT "Molecular cloning, purification, and serological characterization of
RT MP63, a novel antigen secreted by Mycobacterium tuberculosis.";
RL Infect. Immun. 65:16-23(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=Erdmann;
RX MEDLINE=97313166; PubMed=9169770;
RA Harth G., Lee B.Y., Horwitz M.A.;
RT "High-level heterologous expression and secretion in rapidly growing
RT nonpathogenic mycobacteria of four major Mycobacterium tuberculosis
RT extracellular proteins considered to be leading vaccine candidates
RT and drug targets ";
RL Infect. Immun. 65:2321-2328(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,
RA Horsby T., Jagels K., Krogh A., Molean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX Pletschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Petersen J., Debby R., Dodson R., Gwyn M.L., Haft D., Hickey E.,
RA DeLong J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Kohnen A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE OF 30-49.
RC SPECIES=M.tuberculosis;

```

```

RX MEDLINE-95183499; PubMed-7878014;
RA Horwitz M.A., Lee B.W., Dillon B.J., Harth G.;
RT "Protective immunity against tuberculosis induced by vaccination with
RT major extracellular proteins of Mycobacterium tuberculosis.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:1530-1534(1995).
RN [6]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC SPECIES=M.tuberculosis;
RA MEDLINE-91099899; PubMed-1898899;
RA Nagai S., Wiker H.G., Harboe M., Kinomoto M.;
RT "Isolation and partial characterization of major protein antigens in
RT the culture fluid of Mycobacterium tuberculosis.";
RL Infect. Immun. 59:372-382(1991).
RN [7]
RP SEQUENCE FROM N.A.
RC SPECIES=M.dovisi; STRAIN=BCG / Tokyo;
RA Kamile K., Matsuda S., Kobayashi A., Kobayashi K.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Secreted.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U27119; AAB95083.1; -
DR EMBL: U82234; AAB61537.1; -
DR EMBL: Z84498; CAB06500.1; -
DR EMBL: AE007052; AAK46249.1; -
DR EMBL: AB048799; BAB39210.1; -
DR TIGR: MT1977; -
DR TubercuList: RV1926c; -
KW Antigen; Signal; Complete proteome.
FT SIGNAL 1 29
FT CHAIN 30 159
SQ SEQUENCE 159 AA; 16514 MW; EB563E1EABCE428 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 159;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 PVALAAY 20
DB 25 PVALAAY 31

RESULT 14
ID GRAL_TOXGO STANDARD; PRT; 190 AA.
AC P13403;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dense granule protein 1 precursor (Protein GRA 1) (Major antigen
DE P24).
DE GN GRAL.
OS Toxoplasma gondii.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;
OC Toxoplasma.
OX NCB1_Taxid=5811;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-90017513; PubMed-2798425;
RA Cesbron-Delauw M.F., Guy B., Torpier G., Pierce R.J., Lenzen G.,
RA Cesbron J.Y., Charif H., Lepage P., Darcy F., Lecocq J.P.,
RA Capron A.;
RT "Molecular characterization of a 23-kilodalton major antigen secreted
RT by Toxoplasma gondii.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7537-7541(1989).
CC -1- SUBCELLULAR LOCATION: LOCATED IN DENSE GRANULES OF TACHYZOITES.

```

```

CC -1- CAUTION: ACCORDING TO REF.1 THERE ARE TWO POTENTIAL EF-HAND
CC REGIONS. BUT THE EVIDENCE SEEMS WEAK.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M26007; AAA30141.1; -
DR PIR: A33839; A33839.
KW Antigen; Signal.
FT SIGNAL 1 24
FT CHAIN 25 190
FT CARBHYD 30 30
SQ SEQUENCE 190 AA; 20149 MW; 3B05D2610C615A53 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 190;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 529 DAVSAVA 535
DB 149 DAVSAVA 155

RESULT 15
ID RS7_AVIMR STANDARD; PRT; 190 AA.
AC Q92NS1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 40S ribosomal protein S7.
DE GN RS7.
OS Avicennia marina.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Avicenniaceae; Avicennia.
OX NCB1_Taxid=82927;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RA Parani M., Senthilkumar P., Lakshmi M., Parida A.;
RT "A nuclear gene from Avicennia marina homologous to the 40S ribosomal
RT protein S7 gene of animals and human.";
RL (in) Plant Gene Register PGR8-203.
CC -1- SIMILARITY: BELONGS TO THE S7E FAMILY OF RIBOSOMAL PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF056316; AAD03501.1; -
DR EMBL: AF098519; AAC97947.1; -
DR InterPro: IPR000554; Ribosomal_S7E.
DR Pfam: PF01251; Ribosomal_S7e; 1.
DR ProDom: PD006276; Ribosomal_S7e; 1.
DR PROSITE: PS00948; RIBOSOMAL_S7E; FALSE-NEG.
KW Ribosomal protein.
SQ SEQUENCE 190 AA; 21871 MW; CA5173ACF57A47B0 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 190;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 804 HVPYRLR 810

```

DB 64 HYPERLR 70

|||||||

RESULT 16  
ID INO1\_BOVIN STANDARD: PRT: 195 AA.  
AC P07352:  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Interferon omega-1 precursor (Interferon alpha-II-1) (IFN-omega-c1).  
GN IFNM.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=85187974; PubMed=2985969;  
RA Capon D.J., Shepard H.M., Goeddel D.V.;  
RT "Two distinct families of human and bovine interferon-alpha genes are  
RT coordinately expressed and encode functional polypeptides."  
RL Mol. Cell. Biol. 5:768-779(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Chung Y.G., Seidel G.E. Jr.;  
RT "Cloning bovine interferon-tau genes and characterizing their  
RT transcriptional expression during early pregnancy."  
RT Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: BELONGS TO THE INTERFERON ALPHA, BETA AND DELTA  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M11002; AAA30578.1; -;  
DR EMBL: AF238610; AAC14167.1; -;  
DR PIR: B23285; IYBOIT.  
DR HSSP: P01563; ZHIE.  
DR InterPro: IPR000471; Interferon\_abd.  
DR Pfam: PF00143; Interferon\_1.  
DR PRINTS: PR00266; INTERFERONAB.  
DR ProDom: PD000550; Interferon\_abd\_1.  
DR SMART: SM00076; IFabd\_1.  
DR PROSITE: PS00252; INTERFERON\_A-B\_D; 1.  
KW Cytokine; Antiviral; Multigene family; Signal.  
FT SIGNAL 1 23  
FT CHAIN 1 195 INTERFERON OMEGA-1.  
FT DISULFID 24 122 BY SIMILARITY.  
FT DISULFID 52 162 BY SIMILARITY.  
SO SQUENCE 195 AA; 21980 MW; 1B65FD4HCABB2A15 CRC64;  
Query Match 0.8%; Score 7; DB 1; Length 195;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 389 LEQOLL 395  
DB 104 LEQOLL 110  
RESULT 17  
FDOI\_ECOLI STANDARD: PRT: 211 AA.  
ID FDOI\_ECOLI

AC P32174;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Formate dehydrogenase, cytochrome b556(FDO) subunit (Formate  
DE dehydrogenase-O gamma subunit) (FDO-Z gamma subunit) (Aerobic formate  
DE dehydrogenase cytochrome b556 subunit).  
GN FDOI OR B3892 OR Z5434 OR ECS4818.  
OS Escherichia coli, and  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562, 83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE=93347969; PubMed=8346018;  
RA Plunkett G., III, Burland V.D., Daniels D.L., Blattner F.R.;  
RT "Analysis of the Escherichia coli genome. III. DNA sequence of the  
RT region from 87.2 to 89.2 minutes."  
RL Nucleic Acids Res. 21:3391-3398(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
RA Apodaca J., Mantharman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."  
RT Nature 409:529-533(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-O157:H7 / RIMD 0509952;  
RX MEDLINE=21156331; PubMed=11238796;  
RA Hayashi T., Makino K., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kudara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12."  
RL DNA Res. 8:11-22(2001).  
RN [4]  
RP CHARACTERIZATION.  
RC STRAIN-K12;  
RX MEDLINE=96099298; PubMed=8522521;  
RA Abalou H., Pommer J., Giordano G., Mandrand-Berthelet M.-A.;  
RT "Expression and characterization of the Escherichia coli fdo locus  
RT and a possible physiological role for aerobic formate  
RT dehydrogenase."  
RL J. Bacteriol. 177:7141-7149(1995).  
RN [5]  
RP TOPOLOGY.  
RX MEDLINE=99069338; PubMed=9852007;  
RA Benoit S., Abalou H., Mandrand-Berthelet M.-A.;  
RT "Topological analysis of the aerobic membrane-bound formate  
RT dehydrogenase of Escherichia coli."  
RL J. Bacteriol. 180:6625-6634(1998).  
CC -1- FUNCTION: ALLOWS TO USE FORMATE AS MAJOR ELECTRON DONOR DURING  
CC AEROBIC RESPIRATION. SUBUNIT GAMMA IS PROBABLY THE CYTOCHROME  
CC B556(FDO) COMPONENT OF THE FORMATE DEHYDROGENASE.  
CC -1- SUBUNIT: FORMATE DEHYDROGENASE IS A MEMBRANE-BOUND COMPLEX, FORMED  
CC BY SUBUNITS ALPHA, BETA AND GAMMA.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: STRONG, TO PDN1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial



CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; L19201; AAB03025.1; -

DR EMBL; AE000464; AAD13454.1; -

DR EMBL; AE005620; AAG59085.1; -

DR EMBL; AP002567; BAB38241.1; -

DR PIR; S40836; S40836.

KW Ecogene; EG11856; fdoi.

DR Electron transport; Heme; Iron; Transmembrane; Complete proteome.

FT DOMAIN 1 17 CYTOPLASMIC (PROBABLE).

FT TRANSMEM 18 32 PROBABLE.

FT DOMAIN 33 53 PERIPLASMIC (PROBABLE).

FT TRANSMEM 54 72 PROBABLE.

FT DOMAIN 73 112 CYTOPLASMIC (PROBABLE).

FT TRANSMEM 113 130 PROBABLE.

FT DOMAIN 131 151 PERIPLASMIC (PROBABLE).

FT TRANSMEM 152 170 PROBABLE.

FT DOMAIN 171 211 CYTOPLASMIC (PROBABLE).

SQ SEQUENCE 211 AA; 24606 MW; 166AC5A661C738D5 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 211;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 HNLINRD 611  
Db 78 HNLINRD 84  
|||||

RESULT 18  
FIXJ\_AZOCA STANDARD; PRT; 211 AA.  
AC P26487;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Transcriptional regulatory protein fixJ.

GN FIXJ.

OS Azorhizobium caulinodans.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Hyptomicrobium group; Azorhizobium.

OX NCBI\_TaxID=7;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ORS571;

RX MEDLINE=91260451; PubMed=2046550;

RA Kaminski P.A., Elmerich C.;

RT \*Involvement of fixJ in the regulation of nitrogen fixation in Azorhizobium caulinodans.\*;

RL Mol. Microbiol. 5:665-673(1991).

CC -!- FUNCTION: FIXJ, WHEN ACTIVATED BY FIXL, INDUCES THE EXPRESSION OF BOTH NIF, REQUIRED FOR ACTIVATION OF CLASSICAL NIF AND FIX GENES, AND FIXK, REQUIRED FOR FIXN ACTIVATION.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -!- PTM: PHOSPHORYLATED BY FIXL (PROBABLE).

CC -!- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.

CC -!- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL REGULATORS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X56658; CAA39980.1; -

DR PIR; S15167; S15167.

DR HSSP; P10958; IDBM.

DR InterPro; IPR000792; HTH\_LuxR.

DR InterPro; IPR001789; Response\_reg.

DR Pfam; PF00072; response\_reg; 1.

DR Pfam; PF00196; GerE; 1.

DR PRINTS; PR00038; HTHLUXR.

DR ProDom; PD000039; Response\_reg; 1.

DR ProDom; PD000307; HTH\_LuxR; 1.

DR SMART; SM00421; HTH\_LUXR; 1.

DR SMART; SM00448; REC; 1.

DR PROSITE; PS00622; HTH\_LUXR\_FAMILY; 1.

DR PROSITE; PS0110; RESPONSE REGULATORY; 1.

KW Sensory transduction; Phosphorylation; Transcription regulation; DNA-binding; Activator; Nitrogen fixation

FT DOMAIN 6 120 RESPONSE REGULATORY.

FT MOD\_RES 55 55 PHOSPHORYLATION (BY SIMILARITY).

FT DNA\_BIND 160 179 H-T-H MOTIF (BY SIMILARITY).

SQ SEQUENCE 211 AA; 22282 MW; E0F06B3EA4CD827D CRC64;

Query Match 0.8%; Score 7; DB 1; Length 211;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 ALLAGVA 727  
Db 197 ALLAGVA 203  
|||||

RESULT 19  
VARD\_THERH STANDARD; PRT; 223 AA.  
ID VARD\_THERH  
AC 087880;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE V-type ATP synthase subunit D (BC 3.6.3.14) (V-type ATPase subunit D).

GN ATPD OR VARD.

GN Thermus thermophilus.

OS Thermus thermophilus.

OC Bacteria; Thermus/Delinococcus group; Deinococci; Thermales;

OC Thermaceae; Thermus.

OX NCBI\_TaxID=274;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=HB8 / ATCC 27634;

RX MEDLINE=20250964; PubMed=10788522;

RA Yokoyama K., Onkuma S., Taguchi H., Yasunaga T., Wakabayashi T., Yoshida M.;

RT "V-type H<sup>+</sup>-ATPase/synthase from a thermophilic eubacterium, Thermus thermophilus. Subunit structure and operon.";

RL J. Biol. Chem. 275:13955-13961(2000).

CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON GRADIENT ACROSS THE MEMBRANE.

CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate + H(+) (Out).

CC -!- SIMILARITY: BELONGS TO THE V-ATPASE D SUBUNIT FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; D63799; BAA33198.1; ALT\_INIT.

DR InterPro; IPR002699; ATPsynT\_Dsub.

DR Pfam; PF01813; ATP-synt\_D; 1.

DR ProDom; PD004122; ATPsynT\_Dsub; 1.

DR TIGRPFAM; TIGR00309; V\_ATPase\_subD; 1.

KW Hydrolase; ATP synthetase; Hydrogen ion transport.

SQ SEQUENCE 223 AA; 24677 MW; B66AE00DFE0B08B CRC64;

Query Match 0.8%; Score 7; DB 1; Length 223;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 179 KVPRLKA 185  
 DB 100 KVPRLKA 106

## RESULT 20

GPH\_VIBCH STANDARD: PRT: 226 AA.  
 ID GPH\_VIBCH  
 AC 09KNV6;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phosphoglycolate phosphatase (EC 3.1.3.18) (PGP).  
 GN GPH OR VC2624.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / Serotype O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nieman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae.";  
 RL Nature 406:477-483(2000).  
 CC -1- CATALYTIC ACTIVITY: 2-phosphoglycolate + H(2)O = glycolate +  
 CC phosphate.  
 CC -1- SIMILARITY: BELONGS TO THE CBBY/CBBZ/GPH/YIEH FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE004329; AAF5765.1; -  
 CC TIGR: VC2624; -  
 CC DR InterPro: IPR001454; Hlgase/hydrlase.  
 CC DR Pfam: PF00702; Hydrolase; 1.  
 CC KM Carboxylate metabolism; Hydrolase; Complete proteome.  
 CC SQ SEQUENCE 226 AA; 24560 MW; 83F95B10DBF53CB3 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 226;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 241 AARAVGY 247  
 DB 29 AARAVGY 35

RESULT 21  
 COX2\_ASCSU STANDARD: PRT: 232 AA.  
 ID COX2\_ASCSU  
 AC P24882;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1).  
 GN COII.  
 OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).  
 OC Mitochondrion  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;  
 OC Ascarididae; Ascaris.  
 OX NCBI\_TaxID=6253;

RESULT 22  
 RSTA\_ECOLI STANDARD: PRT: 242 AA.  
 ID RSTA\_ECOLI  
 AC P52106;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Transcriptional regulatory protein rsta.  
 GN RSTA OR URPT OR B1608.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Body wall muscle, and Egg;  
 RX MEDLINE=92201635; PubMed=1551572;  
 RA Okimoto R., Macfarlane J.L., Clary D.O., Wolstenholme D.R.;  
 RT "The mitochondrial genomes of two nematodes, Caenorhabditis elegans  
 and Ascaris suum";  
 RL Genetics 130:471-498(1992).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
 CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.  
 CC -1- CATALYTIC ACTIVITY: 4 ferriocytochrome c + O(2) = 4 ferriocytochrome  
 CC c + 2 H(2)O.  
 CC -1- COFACTOR: Copper A.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 CC inner membrane.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X54253; CAA38172.1; -  
 CC PIR: S26023; S26023.  
 CC DR InterPro: IPR001505; Copper\_CuA.  
 CC DR InterPro: IPR002429; Cyl\_c-ox\_2.  
 CC DR Pfam: PF00116; COX2; 1.  
 CC DR Pfam: PF02790; COX2\_TM; 1.  
 CC DR ProDom: PD000131; Copper\_CuA; 1.  
 CC DR PROSITE: PS00078; COX2; 1.  
 CC KM Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;  
 CC Electron transport; Respiratory chain.  
 CC KW DOXAIN 1  
 CC FT TRANSMM 31 52 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  
 CC FT DOXAIN 53 69 MITOCHONDRIAL MATRIX (POTENTIAL).  
 CC FT TRANSMM 70 89 POTENTIAL.  
 CC FT DOXAIN 90 232 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  
 CC FT METAL 164 164 COPPER A (PROBABLE).  
 CC FT METAL 199 199 COPPER A (PROBABLE).  
 CC FT METAL 203 203 COPPER A (PROBABLE).  
 CC FT METAL 207 207 COPPER A (PROBABLE).  
 CC SQ SEQUENCE 232 AA; 26517 MW; 58F9B683FF9DA4C5 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 232;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 SFMPVAL 17  
 DB 208 SFMPVAL 214

RESULT 22  
 RSTA\_ECOLI STANDARD: PRT: 242 AA.  
 ID RSTA\_ECOLI  
 AC P52106;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Transcriptional regulatory protein rsta.  
 GN RSTA OR URPT OR B1608.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-K12;  
 RA Kuempel P.L.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE-97426617; PubMed-9278503;  
 RA Blatner F.N., Plunkett G., Ilt, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-97251357; PubMed-9097039;  
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,  
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,  
 RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,  
 RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubaram S.,  
 RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,  
 RA Yamamoto Y., Horuchi T.;  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 28.0-40.1 min region on the linkage map.";  
 RL DNA Res. 3:363-377(1996).  
 RN [4]  
 RP MAPING.  
 RA MEDLINE-92356829; PubMed-1495392;  
 RX Roedelstein B.A., Kuempel P.L.;  
 RT "In vivo characterization of tus gene expression in Escherichia  
 coli.";  
 RL Mol. Microbiol. 6:1655-1661(1992).  
 CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM RSTB/RSTA.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- PTM: PHOSPHORYLATED BY RSTB (POTENTIAL).  
 CC -1- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U41101; AAA82081.1; -;  
 DR EMBL: AE000256; AAC74680.1; -;  
 DR EMBL: D90803; BAA15346.1; -;  
 DR EMBL: D90804; BAA15356.1; -;  
 DR HSSP: P03025; IOPC.  
 DR Ecogene: EG13190; rsta.  
 DR InterPro: IPR001789; Response\_reg.  
 DR InterPro: IPR001867; Trans\_reg.C.  
 DR Pfam: PF00072; response\_reg.1.  
 DR Pfam: PF00486; trans\_reg.C.1.  
 DR ProDom: PD000039; Response\_reg.1.  
 DR ProDom: PD000329; Trans\_reg.C.1.  
 DR SMART: SM00448; REC.1.  
 DR PROSITE: PS50110; RESPONSE\_REGULATORY.1.  
 DR Sensory transduction; Phosphorylation; Transcription regulation;  
 KW DNA-binding; Complete proteome.  
 FT DOMAIN 6 119 RESPONSE REGULATORY.  
 FT MOD\_RES 55 55 PHOSPHORYLATION (BY SIMILARITY).  
 SO SEQUENCE 242 AA; 27048 MW; 8B5A3068829811B6 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 242;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 334 LDRSDVY 340  
 |||||

Db 201 LDRSDVY 207  
 RESULT 23  
 CIDI\_HUMAN  
 ID CIDI\_HUMAN STANDARD: PRT: 261 AA.  
 AC P56856; Q36PH4;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Claudin-18.  
 CN CLDN18.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE-21470315; PubMed-11585919;  
 RA Nishi T., Nagashima K., Ward J.M., Minoo P., Zimonjic D.B.,  
 RA Popescu N.C., Kimura S.;  
 RT "Claudin-18, a novel downstream target gene for the T/EBP/NKX2.1  
 homeodomain transcription factor, encodes lung- and stomach-specific  
 isoforms through alternative splicing.";  
 RL Mol. Cell. Biol. 21:7380-7390(2001).  
 CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: A1 (shown here) and A2; are  
 CC produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Concentrated at the cell-cell borders of  
 CC epithelial cells.  
 CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF221069; AF26448.1; -;  
 DR EMBL: AF349452; AAL15637.1; -;  
 DR Genew: HGNC:2039; CLDN18.  
 DR InterPro: IPR001832; Claudin.  
 DR InterPro: IPR004031; PMP22.Claudin.  
 DR Pfam: PF00822; PMP22.Claudin; 1.  
 DR PROSITE: PS01346; CLAUDIN; 1.  
 KW Tight junction; Transmembrane; Alternative splicing.  
 FT TRANSMEM 7 27  
 FT TRANSMEM 81 101  
 FT TRANSMEM 123 143  
 FT TRANSMEM 175 195  
 FT VARSPLIC 1 69  
 FT MSTRYCVVAFLSLIAGCIATGDMSTODLYDNPVT  
 FT SVFYEGLMRSCVSSGFTFCRPFTI -> MAVTACGL  
 FT GFVVLIGIAGIATGDMSTODLYDNPVTAVFNVQGLW  
 FT RSCVRESSGFTFCRPFTI (IN ISOFORM A2).  
 SO SEQUENCE 261 AA; 27856 MW; 4362B590D3C8B387 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 404 LQAVRAL 410  
 |||||  
 Db 76 LQAVRAL 82

RESULT 24  
 CIDI\_MOUSE  
 ID CIDI\_MOUSE STANDARD: PRT: 264 AA.  
 AC P56857; Q91ZL1; Q91Z20; Q91Z99;  
 DT 30-MAY-2000 (Rel. 39, Created)

```

DT 30-MAY-2000 (Rel. 39, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE CLAUDIN-18.
GN CLDN18.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=21470315; PubMed=11585919;
RA Nihl T., Nagashima K., Ward J.M., Munoo P., Zimonjic D.B.,
RA Popescu N.C., Kimura S.,
RT "Claudin-18, a novel downstream target gene for the T/EBP/NKX2.1
RT homeodomain transcription factor, encodes lung- and stomach-specific
RT isoforms through alternative splicing."
RL Mol. Cell. Biol. 21:7380-7390(2001).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; A1.1 (shown here), A1.2, A2.1
CC and A2.2: are produced by alternative splicing.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF221068; AAF26447.1; -
DR EMBL: AF349450; AAL15635.1; -
DR EMBL: AF349451; AAL15636.1; -
DR EMBL: AF349453; AAL15638.1; -
DR MGD: MGI:1929209; Clnd18.
DR InterPro: IPR001832; Claudin.
DR Pfam: PF00822; PMP22.Claudin.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane; Alternative splicing.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT VARSPLIC 1 69 MATTCOVGGLLSLGIACIAATGMDMSTODLYDNPT
FT VARSPLIC 1 69 AVFOYEGLMRSCVOSSGFTECRPYFI -> MSYTAQGL
FT VARSPLIC 1 69 GFVNSLIGFAGIIATCNDMDSTODLYNNPTAVFNVOGLW
FT VARSPLIC 1 69 RSCVRESGFEFCRGFTL (IN ISOFORM A2.1 AND
FT VARSPLIC 1 69 ISOFORM A2.2).
FT VARSPLIC 1 69 N -> K (IN ISOFORM A1.2 AND ISOFORM
FT VARSPLIC 1 69 A2.2).
FT VARSPLIC 1 69 MISSING (IN ISOFORM A1.2 AND ISOFORM
FT VARSPLIC 1 69 A2.2).
SQ SEQUENCE 264 AA: 28122 MW: 3CA0D441C4705653 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 264;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 404 LQAVRAL 410
DB 76 LQAVRAL 82

```

```

RESULT 25
PURR_LACLA STANDARD; PRT; 271 AA.
AC 053065;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pur operon repressor.

```

```

GN PURR OR LL2259.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis), and
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360, 1359;
RN [1]
RP SEQUENCE FROM N.A.
RX SPECIES=L.1.lactis; STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Mincker P., Manger S., Jallion O., Malarme K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.,
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=L.1.cremoris; STRAIN=MG1363;
RA Kilstrup M.;
RL Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CONTROLS TRANSCRIPTION OF THE PUR OPERON FOR PURINE
CC BIOSYNTHETIC GENES (BY similarity).
CC -1- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE
CC PHOSPHORIBOSYLTRANSFERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AJ222642; CAA10902.1; -
DR EMBL: AE006455; AAK06357.1; -
DR InterPro: IPR000836; PRTtransferase.
DR InterPro: IPR002375; Pr/PY-TP-Transf.
DR Pfam: PF00156; Pribosyltran; 1.
DR PROSITE: PS00103; PUR_PYR_PR_TRANSFER; FALSE_NEG.
KW DNA-binding; Transcription regulation; Repressor; Complete proteome.
SQ SEQUENCE 271 AA: 30361 MW: 04614AA24E1C4BCD CRC64;

```

```

Query Match 0.8%; Score 7; DB 1; Length 271;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 722 LLAGYAV 728
DB 223 LLAGYAV 229

RESULT 26
LIGE_PSEPA STANDARD; PRT; 280 AA.
ID LIGE_PSEPA
AC P27457;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Beta-etherase (beta-aryl ether cleaving enzyme).
GN LIGE.
OS Pseudomonas paucimobilis (Sphingomonas paucimobilis).
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Sphingomonas.
OX NCBI_TaxID=13689;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RX STRAIN=SYK-6;
RX MEDLINE=92078104; PubMed=1744051;
RA Masai E., Katayama Y., Kawai S., Nishikawa S., Yamasaki M.,
RA Morohoshi N.;
RT "Cloning and sequencing of the gene for a pseudomonas paucimobilis
RT enzyme that cleaves beta-aryl ether."
RL J. Bacteriol. 173:7950-7955(1991).
CC -1- FUNCTION: ABLE TO DEGRADE VARIOUS DIMERIC LIGNIN COMPOUNDS.

```

RA	Azavedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA	Borries R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
RA	Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA	Choi S.K., Codani J.J., Conerton I.F., Cumings N.J., Daniel R.A.,
RA	Denizot F., Devine K.M., Dusterhoft A., Ehlich S.D., Emerson P.T.,
RA	Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Gallero N.,
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA	Gutseppl G., Guy B.J., Haga K., Haleh J., Harwood C.R., Hehaut A.,
RA	Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA	Kobayashi Y., Koester P., Konigstein G., Krogh S., Kumano M.,
RA	Kunita K., Lapidos A., Lardinois S., Lauber J., Lazarevic V.,
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA	Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Nodack M.,
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA	Parro V., Pohl T.M., Portetelle D., Porvollik S., Prescott A.M.,
RA	Prescott E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA	Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
RA	Sorokin A., Taccon E., Takagi T., Takahashi H., Takemaru K.,
RA	Takenuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA	Tosato V., Uchiyama S., Vandebol M., Vannier F., Vassartoli A.,
RA	Viati A., Wambut R., Wedler E., Wedler H., Weltzenegger T.,
RA	Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA	Yoshida K., Yoshikawa H.F., Zumsstein E., Yoshikawa H., Danchin A.,
RT	"The complete genome sequence of the Gram-positive bacterium <i>Bacillus</i>
RT	<i>subtilis</i> ".
RL	Nature 390:249-256(1997).
CC	-i- SOURCE/LOC: LOCATION: Integral membrane protein (Potential).
CC	-i- SIMILARITY: TO B.SUBTILIS YPIC AND YQFU.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL: D16311; BAO03818.1; -
DR	EMBL: Y09476; CAA70630.1; -
DR	EMBL: Z79580; CAB01834.1; -
DR	EMBL: Z99109; CAB12952.1; -
DR	PIR: A49921; A49921.
DR	Subtilist: BG10850; Y1FT.
DR	InterPro: IPR003740; DUF161.
DR	Pfam: PF02588; DUF161; 1.
KM	Hypothetical protein; Transmembrane; Complete proteome.
FT	TRANSMEM 9 29 POTENTIAL.
FT	TRANSMEM 54 74 POTENTIAL.
FT	TRANSMEM 80 100 POTENTIAL.
FT	TRANSMEM 151 171 POTENTIAL.
FT	CONFLICT 121 121 V -> D (IN REF. 1).
FT	CONFLICT 226 226 V -> L (IN REF. 1).
FT	CONFLICT 278 278 R -> T (IN REF. 1).
SO	SEQUENCE 280 AA; 30535 MW; 4BF98B1288DA59E7 CNC64;
QY	70 LNAAGLN 76
Db	18 LNAAGLN 24
RESULT 28	
THTR_HUMAN	
ID THTR_HUMAN	STANDARD: PRT; 296 AA.
AC Q16762;	
DT 15-JUL-1998 (Rel. 36, Created)	

DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Thiosulfate sulfotransferase (EC 2.8.1.1) (Rhodanese).  
 GN TST.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fetal Liver;  
 RX MEDLINE=97223396; PubMed=9070219;  
 RA Alta N., Ishii K., Akamatsu Y., Ogasawara Y., Tanabe S.;  
 RT Cloning and expression of human liver rhodanese cDNA.";  
 RL Biochem. Biophys. Res. Commun. 231:56-60(1997).  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057165; PubMed=10591208;  
 RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,  
 RA Clamp M., Slink L.J., Alnough R., Almeida J.P., Habbage A.K.,  
 RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,  
 RA Burdill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
 RA Clegg S.M., Cobley V.E., Cole G.G., Collier R.E., Connor R.,  
 RA Conroy D., Corry N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
 RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
 RA Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,  
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,  
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,  
 RA Lait D.G., Langford C.F., Leverisha M.A., Lloyd C., Lloyd D.M.,  
 RA McLay J., McLaren S., McMurphy A.A., Matthews L.H., Mccann O.T.,  
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
 RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,  
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,  
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
 RA Miosshama S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuama S.,  
 RA Roe B.A., Chen F., Chu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.T.,  
 RA Dorman A., Fang F., Fu Y., Gribble J., Deschamps S., Do A., Do T.,  
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shauli S., Sloan D., Song L.,  
 RA Wang O., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
 RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,  
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozeresky P., Rohlfing T.,  
 RA Schaefer P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,  
 RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,  
 RA Emanuel B.S., Shaikh T., Kuranashi H., Salta S., Budarf M.L.,  
 RA Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,  
 RA Kim U.-J., Shizuya H., Simon M.I., Dumanaki J.P., Peyard K.P., Kedra D.,  
 RA Seroussi E., Franssen I., Tapia I., Bruder C.E., O'Brien K.P.,  
 RA Wilkinson P., Podetlech A., Hartman K., Hu X., Khan A.S., Lane L.,  
 RA Tlahun Y., Wright H.;  
 RT "The DNA sequence of human chromosome 22.";  
 RL Nature 402:489-495(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: FORMATION OF IRON-SULFUR COMPLEXES. CYANIDE  
 CC DETOXIFICATION OR MODIFICATION OF SULFUR-CONTAINING ENZYMES.  
 CC OTHER THIOL COMPOUNDS, BESIDES CYANIDE, CAN ACT AS SULFUR ION  
 CC ACCEPTORS. ALSO HAS WEAK MERCAPTOPROPYLATE SULFOTRANSFERASE (MST)  
 CC ACTIVITY (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.  
 CC -!- SUBUNIT: MONOMER.

CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.  
 CC -!- DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR  
 CC CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,  
 CC THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D87292; BAAL3327.1; -;  
 CC EMBL: 273420; CA97762.1; -;  
 CC EMBL: BC010148; AAH01048.1; -;  
 CC HSSP: P00586; IRRS.  
 CC Genew: HGNC:12388; TST.  
 CC MIM: 180370; -;  
 CC InterPro: IPR001763; Rhodanese-like.  
 CC InterPro: IPR001307; Rhodanese.  
 CC Pfam: PF00581; Rhodanese; 2.  
 CC SMART: SM00450; RHOD; 2.  
 CC PROSITE: PS00380; RHODANES\_1; 1.  
 CC PROSITE: PS00683; RHODANES\_2; 1.  
 CC KW Transferase; Mitochondrion.  
 CC FT INIT MET 0 BY SIMILARITY.  
 CC FT DOMAIN 1 142 A DOMAIN.  
 CC FT DOMAIN 143 158 HINGE.  
 CC FT DOMAIN 159 296 B DOMAIN.  
 CC FT ACT\_SITE 186 186 MAY PLAY A ROLE IN SUBSTRATE BINDING (BY  
 CC SIMILARITY).  
 CC FT ACT\_SITE 247 247 SUBSTRATE (THIOSULFATE) BINDING  
 CC FT ACT\_SITE 248 248 (BY SIMILARITY).  
 CC FT ACT\_SITE 249 249 SUBSTRATE (THIOSULFATE) BINDING  
 CC FT ACT\_SITE 249 249 (BY SIMILARITY).  
 CC SQ SEQUENCE 296 AA; 33297 MW; 872C52008A8DC5B CRC64;  
 CC Query Match 0.88; Score 7; DB 1; Length 296;  
 CC Best Local Similarity 100.0%; Pred. No. 62;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC QY 15 VALAAYL 21  
 CC 1111111  
 CC Db 256 VALAAYL 262  
 CC RESULT 29  
 CC THTR\_MOUSE  
 CC ID THTR\_MOUSE STANDARD; PRT; 296 AA.  
 CC AC P52196;  
 CC DT 01-OCT-1996 (Rel. 34, Created)  
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 CC DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 CC DE Thiosulfate sulfotransferase (EC 2.8.1.1) (Rhodanese).  
 CC GN TST.  
 CC OS Mus musculus (Mouse).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC OX NCBI\_TaxID=10090;  
 CC RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BAIB/C; TISSUE=Liver;  
 RX MEDLINE=96074596; PubMed=7488186;  
 RA Dooley T.P., Nair S.K., Garcia R.E., Courtney B.C.;  
 RT "Mouse rhodanese gene (Tst): cDNA cloning, sequencing, and  
 RT recombinant protein expression.";  
 RL Biochem. Biophys. Res. Commun. 216:1101-1109(1995).  
 CC -!- FUNCTION: FORMATION OF IRON-SULFUR COMPLEXES AND CYANIDE  
 CC DETOXIFICATION.  
 CC -!- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.  
 CC -!- SUBUNIT: MONOMER.

```

CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -1- TISSUE SPECIFICITY: FOUND IN NUMEROUS TISSUES.
CC -1- DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR
CC CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,
CC THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.
CC -1- SIMILARITY: BELONGS TO THE RHODANSE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U35741; AAC52342.1; -.
CC HSSP: P00586; 1RHS.
CC SMISS-2DPAGE: P52196; MOUSE.
CC MGD: MGI:98852; Tst.
CC InterPro: IPR001763; Rhodanese-like.
CC InterPro: IPR001307; Rhodanese.
CC Pfam: PF00581; Rhodanese; 2.
CC SMART: SM00450; RHOD; 2.
CC PROSITE: PS00380; RHODANSE_1; 1.
CC PROSITE: PS00683; RHODANSE_2; 1.
CC K1M
CC Transferase: Mitochondrion.
CC INIT_MET 0
CC BY SIMILARITY.
CC FT DOMAIN 1 142 A DOMAIN.
CC FT DOMAIN 143 158 HINGE.
CC FT DOMAIN 159 296 B DOMAIN.
CC FT ACT_SITE 186 MAY PLAY A ROLE IN SUBSTRATE BINDING (BY
CC ACT_SITE 247 SIMILARITY).
CC FT ACT_SITE 248 SUBSTRATE (THIOSULFATE) BINDING
CC FT ACT_SITE 249 (BY SIMILARITY).
CC FT ACT_SITE 249 SUBSTRATE (THIOSULFATE) BINDING
CC FT ACT_SITE 249 (BY SIMILARITY).
CC SQ SEQUENCE 296 AA; 33334 MW; 82089D80F9AE55A CRC64;
Query Match 0.8%; Score 7; DB 1; Length 296;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 15 VALAAYL 21
Db 256 VALAAYL 262
RESULT 30
34KD_MYCTU STANDARD; PRT; 303 AA.
ID 34KD_MYCTU
AC P71556;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 34 kDa antigenic protein homolog.
GN RV0954 OR MT0981 OR MTCY10D7.20C.
OS Mycobacterium tuberculosis.
CC Bacteria: Actinobacteria: Actinobacteria (class): Actinobacteridae;
CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CC NCBI_TaxID=1773;
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruster S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;

```

```

RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Mayhew L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: STRONG, TO M.PARATUBERCULOSIS 34 kDa ANTIGENIC
CC PROTEIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: 279700; CAB01996.1; -.
CC DR EMBL: AE006983; AAK45229.1; -.
CC DR TIGR: MT0981; -.
CC DR Tuberculist; RV0954; -.
CC FT TRANSMEM 42 62 POTENTIAL.
CC FT TRANSMEM 77 97 POTENTIAL.
CC FT TRANSMEM 102 122 POTENTIAL.
CC FT TRANSMEM 134 154 POTENTIAL.
CC SQ SEQUENCE 303 AA; 30204 MW; 4FE18A077FDFDB85 CRC64;
Query Match 0.8%; Score 7; DB 1; Length 303;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 721 ALLAGVA 727
Db 88 ALLAGVA 94
RESULT 31
OYE6_HUMAN STANDARD; PRT; 313 AA.
ID OYE6_HUMAN
AC Q96RD3;
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Olfactory receptor 52B6.
GN OR52B6.
OS Homo sapiens (Human).
CC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi;
CC Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.
CC NCBI_TaxID=9606;
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 70-287 FROM N.A.
RA Fuchs T., Malecova B., Linhart C., Sharan R., Khen M., Herwig R.,
RA Shmulevich D., Elkon R., Steinfeld M., O'Brien J.K., Redelof U.,
RA Lehrach H., Olander Z., Glusman G., Lancet D., Shamir R.;
RT "DEFOG: a practical scheme for deciphering families of genes.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: AF399504; AAK94989.1; -
DR InterPro: IPR000276; GPCR_Rhodopsin.
DR Pfam: PF00001; 7tm.1; 1.
DR PROSITE: PS00237; G_PROTEIN_RECPT_FL1; 1.
DR PROSITE: PS50262; G_PROTEIN_RECPT_FL2; 1.
KW G-protein coupled receptor; Transmembrane; glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 27 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 28 51 1 (POTENTIAL).
FT DOMAIN 52 59 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 60 81 2 (POTENTIAL).
FT DOMAIN 82 102 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 103 122 3 (POTENTIAL).
FT DOMAIN 123 141 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 142 160 4 (POTENTIAL).
FT DOMAIN 161 197 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 198 220 5 (POTENTIAL).
FT DOMAIN 221 237 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 238 260 6 (POTENTIAL).
FT DOMAIN 261 275 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 276 295 7 (POTENTIAL).
FT DOMAIN 296 313 CYTOPLASMIC (POTENTIAL).
FT CAROHPD 5 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 313 AA; 35523 MW; 6FC174DBDC20BB3 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 313;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 VVVPPTL 157
| | | | |
Db 282 VVVPPTL 288

RESULT 32
MRAM_FUSUN
ID MRAM_FUSUN STANDARD; PRT; 314 AA.
AC Q8R6F5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE S-adenosyl-methyltransferase mram (EC 2.1.1.-).
GN MRAM OR FN1711.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapural V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Vasilava O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fongsteil M., Kyriades N., Overbeek R.;
RT Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.
RL J. Bacteriol. 184:2005-2018(2002).
CC -1- FUNCTION: Exhibits a S-adenosyl-dependent methyltransferase
CC activity (by similarity).
CC -1- SIMILARITY: BELONGS TO THE MRAM FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: AE010475; AAL93826.1; -
DR Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 314 AA; 35782 MW; AB1FD9DD3D269136 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 314;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 471 DGLIMDI 477
| | | | |
Db 100 DGLIMDI 106

RESULT 33
MOCA_RHIME
ID MOCA_RHIME STANDARD; PRT; 317 AA.
AC P49307;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Rhizopine catabolism protein moca (EC 1.-.-.-).
GN MOCA.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L5-30;
RX MEDLINE=95147842; PubMed=7845353;
RA Rossbach S., Kulpa D.A., Rossbach U., de Bruijn F.J.;
RT "Molecular and genetic characterization of the rhizopine catabolism
RT (mocaBRC) genes of Rhizobium meliloti L5-30."
RL Mol. Gen. Genet. 245:11-24(1994).
CC -1- FUNCTION: COULD CATALYZE THE NADH-DEPENDENT DEHYDROGENASE REACTION
CC INVOLVED IN RHIZOPINE (L-3-O-METHYL-SCYLLIO-INOSAMINE) CATABOLISM.
CC -1- SIMILARITY: BELONGS TO THE GFO/IDH/MOCA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: X78503; CAA55269.1; -
DR InterPro: IPR000683; GFO_IDH_MOCA.
DR InterPro: IPR004104; GFO_IDH_MOCA_C.
DR Pfam: PF01408; GFO_IDH_MOCA; 1.
KW Pfam; PF02894; GFO_IDH_MOCA_C; 1.
KW Oxidoreductase; NAD.
SQ SEQUENCE 317 AA; 33725 MW; 06CD96EC77276652 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 317;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 752 GSSGLVS 758
| | | | |
Db 220 GSSGLVS 226

RESULT 34
ISPB_ECOLI
ID ISPB_ECOLI STANDARD; PRT; 323 AA.
AC P19641;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Octaprenyl-diphosphate synthase (EC 2.5.1.-) (Octaprenyl pyrophosphate

```



DE synthetase) (opp synthetase).  
 GN ISPB OR CEL OR B3187.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_Taxid=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / W3110;  
 RX MEDLINE=94146411; PubMed=8312607;  
 RA Jeong J.H., Kitakawa M.S., Isono S., Isono K.;  
 RT "Cloning and nucleotide sequencing of the genes, rpu and rpmA, for  
 RL ribosomal proteins L21 and L27 of Escherichia coli.";  
 RN DNA Seq. 4:59-67(1993).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 [3]  
 RP SEQUENCE OF 196-323 FROM N.A.  
 RX MEDLINE=89359178; PubMed=2670911;  
 RA Choi Y.-L., Nishida T., Kawamukai M., Utsumi R., Sakai H., Komano T.;  
 RT "Cloning and sequencing of an Escherichia coli gene, nlp, highly  
 RL homologous to the ner genes of bacteriophages Mu and D108.";  
 RN J. Bacteriol. 171:5222-5225(1989).  
 [4]  
 RN FUNCTION.  
 RP STRAIN-K12 / JM109;  
 RX MEDLINE=94311902; PubMed=8037730;  
 RA Asai K.-I., Fujisaki S., Nishimura Y., Nishino T., Okada K.,  
 RA Natsugawa T., Kawamukai M., Matsuda H.;  
 RT "The identification of Escherichia coli ispb (cel) gene encoding the  
 RL octaprenyl diphosphate synthase.";  
 CC Biochem. Biophys. Res. Commun. 202:340-345(1994).  
 CC -1- FUNCTION: SUPPLIES OCTAPRENYL DIPHOSPHATE, THE PRECURSOR FOR THE  
 CC SIDE CHAIN OF THE ISOPRENOID QUINONES UBIQUINONE AND MENAQUINONE.  
 CC -1- SIMILARITY: BELONGS TO THE PPp/Gpp SYNTHETASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D13267; -; NOT ANNOTATED\_CDS.  
 DR EMBL: U18987; AA57988.1; -;  
 DR EMBL: AE000399; AAC76219.1; -;  
 DR EMBL: X68873; CAA48735.1; ALT\_SEQ.  
 DR PIR: PV0010; PV0010.  
 DR Ecogen: EG10017; ispb.  
 DR Interpro: IPR000092; Polyprenyl-synt.  
 DR Pfam: PF00348; polyprenyl-synt.1.  
 DR PROSITE: PS00444; POLYPRENYL-SYNTHET\_2; 1.  
 DR PROSITE: PS00723; POLYPRENYL-SYNTHET\_1; 1.  
 KW isoprene biosynthesis; Transferase; Complete proteome.  
 SQ SEQUENCE 323 AA; 35217 MW; 08AD7AE7AC230EF8 CRC64;

Query Match 0.8%; Score 7; DB 1; Length 323;  
 Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 241 AARAVGY 247  
 |||||  
 DB 55 AARAVGY 61

RESULT 35  
 E13K\_TOBAC  
 ID E13K\_TOBAC STANDARD; PRT; 331 AA.  
 AC P52398;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Glucan endo-1,3-beta-glucosidase, acidic isoform GL161 precursor  
 DE (EC 3.2.1.39) ((1->3)-beta-glucan endohydrolase) ((1->3)-beta-  
 DE glucanase) (Beta-1,3-endoglucanase).  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_Taxid=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Xanthi NC; TISSUE=Leaf;  
 RA Ward E.R., Payne G.B., Moyer M.B., Williams S.C., Dincher S.S.,  
 RA Sharkey K.C., Beck J.J., Taylor H.T., Ahl-Goy P., Meins F.,  
 RA Ryals J.A.;  
 RT "Differential regulation of beta-1,3-glucanase messenger RNAs in  
 RL response to pathogen infection.";  
 CC Plant Physiol. 96:390-397(1991).  
 CC -1- FUNCTION: IS THOUGHT TO BE AN IMPORTANT PLANT DEFENSE-RELATED  
 CC PRODUCT AGAINST FUNGAL PATHOGENS.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,3-beta-D-glucosidic linkages  
 CC in 1,3-beta-D-glucans.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: IS EXPRESSED PRIMARILY IN EPIDERMAL CELL OF  
 CC HEALTHY PLANT, AND FOLLOWING INDUCTION BY ETHYLENE, ACCUMULATES IN  
 CC MESOPHYLL CELLS.  
 CC -1- INDUCTION: BY VIRAL INFECTION.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 17 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M60464; AAA34053.1; -;  
 DR HSSP: P15737; ICHS.  
 DR Interpro: IPR000490; Glyco\_hydro\_17.  
 DR Pfam: PF00332; Glyco\_hydro\_17; 1.  
 DR PROSITE: PS00587; GLYCOSYL\_HYDROL\_F17; 1.  
 KW Hydrolase; Glycosidase; Signal; Multigene family.  
 FT STGNL 1 26  
 FT CHAIN 27 331  
 FT MOD\_RES 27 27 GLUCAN ENDO-1,3-BETA-GLUCOSIDASE, ACIDIC  
 FT ACT\_SITE 244 244 ISOFORM GL161.  
 FT ACT\_SITE 304 304 PYROLDIONE CARBOXYLIC ACID (PROBABLE).  
 FT CARBOHYD 55 55 NUCLEOPHILE (BY SIMILARITY).  
 FT CARBOHYD 75 75 PROTON DONOR (BY SIMILARITY).  
 FT SEQUENCE 331 AA; 36891 MW; 5F0171DE53451BBA CRC64;  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 Query Match 0.8%; Score 7; DB 1; Length 331;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 345 NTADVSL 351  
 |||||  
 DB 194 NTADVSL 200

RESULT 36  
 HIT1\_HAEIN  
 ID HIT1\_HAEIN STANDARD; PRT; 332 AA.  
 AC P35755; Q53439;

DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Iron-utilization periplasmic protein precursor (Major ferric iron  
 binding protein) (FBP) (Iron-regulated 40 kDa protein) (MIRP) (Fe(+3)-  
 binding protein).  
 GN HTTA OR FBP OR H10097.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 Haemophilus.  
 NCBI\_TaxID:727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fieleschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,  
 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 Venter J.C.;  
 RA \*Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.;  
 RL Science 269:496-512(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-NTHI TN106;  
 RX MEDLINE=95012644; PubMed=7927717;  
 RA Sanders J.D., Cope L.D., Hansen E.J.;  
 RT \*Identification of a locus involved in the utilization of iron by  
 Haemophilus influenzae.;  
 RL Infect. Immun. 62:4515-4525(1994).  
 RN [3]  
 RP SEQUENCE OF 24-45.  
 RC STRAIN-EAGAN / 8358;  
 RX MEDLINE=92210482; PubMed=1556062;  
 RA Hartness R.E., Chong P., Klein M.H.;  
 RT \*Identification of two iron-repressed periplasmic proteins in  
 Haemophilus influenzae.;  
 RL J. Bacteriol. 174:2425-2430(1992).  
 RN [4]  
 RP SEQUENCE OF 24-31.  
 RC MEDLINE=20137488; PubMed=10675023;  
 RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,  
 Gray C., Fountoulakis M.;  
 RT \*Two-dimensional map of the proteome of Haemophilus influenzae.;  
 RL Electrophoresis 21:411-429(2000).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 24-332.  
 RX MEDLINE=98025060; PubMed=9360608;  
 RA Bruns C.M., Nowalk A.J., Arvai A.S., McTigue M.A., Vaughan K.G.,  
 Metzner T.A., McGee D.E.;  
 RT \*Structure of Haemophilus influenzae Fe(+3)-binding protein reveals  
 convergent evolution within a superfamily.;  
 RL Nat. Struct. Biol. 4:919-924(1997).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS) OF 24-332.  
 RX MEDLINE=21614394; PubMed=11747438;  
 RA Bruns C.M., Anderson D.S., Vaughan K.G., Williams P.A., Nowalk A.J.,  
 McGee D.E., Metzner T.A.;  
 RT \*Crystallographic and biochemical analyses of the metal-free  
 Haemophilus influenzae Fe(+3)-binding protein.;  
 RL Biochemistry 40:15631-15637(2001).  
 CC -1- FUNCTION: THIS PROTEIN MAY BE A CENTRAL COMPONENT IN THE IRON-  
 ACQUISITION SYSTEM.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- INDUCTION: BY IRON DEPRIVATION.  
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING  
 CC PROTEIN FAMILY 1.  
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A

CC -----  
 CC FRAMESHIFT IN POSITION 14.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U32695; AAC21773.1; ALT\_FRAME.  
 CC DR EMBL: S72674; AAB32110.1; -.  
 CC DR PIR: A41833; A41833.  
 CC DR PDB: 1MRP; 28-JAN-98.  
 CC DR PDB: 1D9V; 17-NOV-99.  
 CC DR TIGR: H10097; -.  
 CC DR InterPro: IPR000567; SBP\_bac\_1.  
 CC DR Pfam: PF01547; SBP\_BACTERIAL\_1; 1.  
 CC DR PROSITE: PS01037; SBP\_BACTERIAL\_1; 1.  
 CC KW Transport; Iron transport; Periplasmic; Signal; 3D-structure;  
 CC Complete proteome.  
 CC FT SIGNAL 1 23  
 CC FT CHAIN 24 332 IRON-UTILIZATION PERIPLASMIC PROTEIN.  
 CC FT VARIANT 24 24 D -> A (IN STRAINS PAK12085 AND SB33).  
 CC FT VARIANT 37 37 T -> V (IN STRAIN TN106).  
 CC FT VARIANT 37 37 T -> A (IN STRAIN TN106).  
 CC FT VARIANT 59 59 E -> A (IN STRAIN TN106).  
 CC FT VARIANT 103 103 Q -> K (IN STRAIN TN106).  
 CC FT VARIANT 119 119 I -> V (IN STRAIN TN106).  
 CC FT VARIANT 284 284 A -> V (IN STRAIN TN106).  
 CC FT VARIANT 323 323 I -> T (IN STRAIN TN106).  
 CC SQ SEQUENCE 332 AA; 36193 MW; 4E6549FD76262CA6 CRC64;  
 CC  
 CC Query Match 0.8%; Score 7; DB 1; Length 332;  
 CC Best Local Similarity 100.0%; Pred. No. 69;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 238 ALVANA 244  
 CC |||||  
 CC Db 280 ALVANA 286  
 CC  
 CC RESULT 37  
 CC YPFG\_ECOLI  
 CC ID YPFG\_ECOLI STANDARD; PRT; 347 AA.  
 CC AC P76559;  
 CC DT 15-JUL-1998 (Rel. 36, Created)  
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Hypothetical protein ypfG precursor.  
 CC GN YPFG OR B2466.  
 CC OS Escherichia coli.  
 CC OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC OC Escherichia.  
 CC OX NCBI\_TaxID=562;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN-K12 / MG1655;  
 CC RX MEDLINE=97426617; PubMed=9278503;  
 CC RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 CC Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 CC Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 CC Mau B., Shao Y.;  
 CC RT \*The complete genome sequence of Escherichia coli K-12.;  
 CC RL Science 277:12453-12474(1997).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL: AE00333; AAC75519.1; -.
DR Ecogene; EG14194; ypfG.
KM Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 347
SQ SEQUENCE 347 AA; 38746 MW; 1E75666CEB066FF CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 347;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 TP1SLSE 122
Db 193 TP1SLSE 199

RESULT 38
GP21_HUMAN STANDARD; PRT; 349 AA.
ID GP21_HUMAN
AC Q99679;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable G protein-coupled receptor GPR21.
GN GPR21.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-972579; PubMed-9073069;
RA O'Dowd B.F., Nguyen T., Jung B.P., Marchese A., Cheng R.,
RA Heng H.H.O., Kolakowski L.F. Jr., Lynch K.R., George S.R.;
RT "Cloning and chromosomal mapping of four putative human
RT G-protein-coupled receptor genes.";
RT Gene 187:75-81(1997).
RL CC
CC -1- FUNCTION: ORPHAN RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: NOT DETECTED IN THE BRAIN REGIONS THALAMUS,
CC PUTAMEN, CAUDATE, FRONTAL CORTEX, PONS, HYPOTHALAMUS, HIPPOCAMPUS.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U66580; AAC51303.1; -.
DR Genew; HGNC:4476; GPR21.
DR MIM: 601909;
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS50262; G_PROTEIN_RECEP_F1_2; 1.
KM G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 32
FT TRANSMEM 33 53
FT DOMAIN 54 75
FT TRANSMEM 76 96
FT DOMAIN 97 104
FT TRANSMEM 105 125
FT DOMAIN 126 147
FT TRANSMEM 148 168
FT DOMAIN 169 191
FT TRANSMEM 192 212
FT DOMAIN 213 252
FT TRANSMEM 253 273
FT DOMAIN 274 283
EXTRACELLULAR (POTENTIAL).

```

```

FT TRANSMEM 284 304
FT DOMAIN 305 349
FT CARBOHYD 2 2
FT CARBOHYD 8 8
SQ SEQUENCE 349 AA; 39515 MW; 5369DFC16E9EEC4 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 349;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 SSOSGET 196
Db 233 SSOSGET 239

RESULT 39
HRCA_CAUCR STANDARD; PRT; 358 AA.
ID HRCA_CAUCR
AC P54305;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heat-inducible transcription repressor hrca.
GN HRCA OR CC0153.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CB15N / NA1000;
RA MEDLINE-96178944; PubMed-8606155;
RA Roberts R.C., Toochinda C., Avedissian M., Baldini R.L.,
RA Gomes S.L., Shapiro L.;
RT "Identification of a Caulobacter crescentus operon encoding hrca,
RT involved in negatively regulating heat-inducible transcription, and
RT the chapone gene gipe.";
RT J. Bacteriol. 178:1829-1841(1996).
RL CC
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 19089 / CB15;
RA MEDLINE-21173698; PubMed-11259647;
RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uetendack T., Tran K., Wolf A., Yamathavan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -1- FUNCTION: NEGATIVE REGULATOR OF CLASS I HEAT SHOCK GENES (GRPE-
CC DNAB-DNAJ AND GROELS OPERONS). PREVENTS HEAT-SHOCK INDUCTION OF
CC THESE OPERONS.
CC -1- SIMILARITY: BELONGS TO THE HRCA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U33324; AAB01515.1; -.
DR EMBL: AE003689; AAK22140.1; -.
DR TIGR: CC0153;
DR InterPro: IPR002571; Hrca.
DR Pfam: PF01628; Hrca; 1.
DR TIGRGRAMS: TIGR00331; hrca; 1.
KM Transcription regulation; Repressor; Heat shock; Complete proteome.
FT CONFLICT 276 279
SQ SEQUENCE 358 AA; 38265 MW; 38D7D61D943B13C CRC64;

```

Query Match 0.8%; Score 7; DB 1; Length 358;  
 Best Local Similarity 100.0%; Pred. No. 74;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 838 GGOVLAV 844  
 DB 158 GGOVLAV 164

RESULT 40  
 PCE\_TACTR  
 ID PCE\_TACTR STANDARD; PRT: 375 AA.  
 AC P21902;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Proclotting enzyme precursor (EC 3.4.21.86).  
 OS Tachypleus tridentatus (Japanese horseshoe crab).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;  
 OC Limulidae; Tachypleus.  
 OX NCBI\_TaxID=6853;  
 RN [1]  
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RC TISSUE=Hemocyte;  
 RX MEDLINE=91093088; PubMed=2266134;  
 RA Muta T., Hashimoto R., Miyata T., Nishimura H., Toh Y., Iwanaga S.;  
 RT "Proclotting enzyme from horseshoe crab hemocytes. cDNA cloning,  
 disulfide locations, and subcellular localization.";  
 RL J. Biol. Chem. 265:22426-22433(1990).  
 CC -1- FUNCTION: THIS ENZYME IS CLOSELY ASSOCIATED WITH AN ENDOTOXIN-  
 SENSITIVE HEMOLYMPH COAGULATION SYSTEM IN LIMULUS. ITS ACTIVE FORM  
 CATALYZES THE CONVERSION OF COAGULOGEN TO INSOLUBLE COAGULIN GEL.  
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of 18-Arg-1- and 47-Arg-1-  
 bonds in coagulogen to form coagulin and fragments.  
 CC -1- SUBUNIT: UPON ACTIVATION BY FACTOR B, IT IS CONVERTED TO A TWO-  
 CHAIN ACTIVE FORM COMPOSED OF A LIGHT AND A HEAVY CHAIN.  
 CC -1- PPM: CONTAINS SIX O-LINKED CARBOHYDRATE CHAINS IN THE N-TERMINAL  
 LIGHT CHAIN GENERATED AFTER ACTIVATION.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M58366; AAA30094.1; -;  
 DR PIR: A23689; A23689.  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.221; -;  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; Trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRYD\_SPC; 1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolyase; Serine protease; Zymogen; Signal; Hemolymph clotting;  
 KW Glycoprotein.  
 FT SIGNAL 1 17  
 FT PROPEP 18 29 OR 25 (POTENTIAL).  
 FT CHAIN 30 127 PROCLOTTING ENZYME LIGHT CHAIN.  
 FT CHAIN 128 375 PROCLOTTING ENZYME HEAVY CHAIN.  
 FT DOMAIN 128 375 SERINE PROTEASE.  
 FT MOD\_RES 30 30 PYRROLIDONE CARBOXYLIC ACID.  
 FT ACT\_SITE 172 172 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 228 228 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 326 326 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 40 83

FT DISULFID 50 73  
 FT DISULFID 56 84  
 FT DISULFID 118 248 INTERCHAIN.  
 FT DISULFID 157 173  
 FT DISULFID 295 311  
 FT DISULFID 322 351  
 FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 235 235 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .).  
 SQ SEQUENCE 375 AA; 41591 MW; 94209E514DFCF8FB CRC64;

Query Match 0.8%; Score 7; DB 1; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 698 TOAPPET 704  
 DB 94 TOAPPET 100

Search completed: April 28, 2003, 16:33:39  
 Job time : 66 secs

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:30:40 ; Search time 101 Seconds  
(without alignments)  
1874.825 Million cell updates/sec

Title: US-09-914-168-2  
Perfect score: 919  
Sequence: 1 MSKPYLFANRSMFVALAAY.....TGKKEGNDPIKILHFTGTFP 919

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: listing first 1000 summaries

Database :

- 1: SP archaea:\*
- 2: SP bacteria:\*
- 3: SP fungi:\*
- 4: SP human:\*
- 5: SP invertebrate:\*
- 6: SP mammal:\*
- 7: SP mhc:\*
- 8: SP organelle:\*
- 9: SP phage:\*
- 10: SP plant:\*
- 11: SP rodent:\*
- 12: SP virus:\*
- 13: SP vertebrate:\*
- 14: SP unclassified:\*
- 15: SP virus:\*
- 16: SP bacteriaph:\*
- 17: SP archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	1.4	578	16	082B95
2	11	1.2	577	16	08XB95
3	10	1.1	579	16	0910U1
4	9	1.0	260	16	084951
5	9	1.0	262	2	09R802
6	9	1.0	586	16	09CK26
7	9	1.0	1085	2	09RPH7
8	8	0.9	174	16	P95086
9	8	0.9	235	16	098UV3
10	8	0.9	319	16	096HK9
11	8	0.9	331	16	0927W8
12	8	0.9	331	16	08Y4C5
13	8	0.9	340	16	09ZBL4
14	8	0.9	379	2	P72239
15	8	0.9	383	16	09KYD1
16	8	0.9	452	12	08V5A8

17	8	0.9	469	16	086499	086499 streptomyce
18	8	0.9	508	5	076799	076799 dtrophilia
19	8	0.9	508	17	08T139	08T139 methanosarc
20	8	0.9	582	16	09KP31	09KP31 vibrio chol
21	8	0.9	731	16	08ZS53	08ZS53 anabaena sp
22	8	0.9	1047	17	09HK17	09HK17 thermoplasm
23	8	0.9	1373	5	09VKR2	09VKR2 dtrophilia
24	8	0.9	1417	10	048697	048697 arabidopsi
25	8	0.8	33	2	045283	045283 bacillus 11
26	7	0.8	62	9	08W6C9	08W6C9 vibrio phag
27	7	0.8	66	10	08W542	08W542 retama raet
28	7	0.8	66	16	09PD36	09PD36 xylella fas
29	7	0.8	72	16	P74009	P74009 synechocyst
30	7	0.8	82	16	08XCJ5	08XCJ5 salmone
31	7	0.8	84	17	08TRE1	08TRE1 methanosarc
32	7	0.8	85	13	09PVI8	09PVI8 gallus gall
33	7	0.8	87	16	09CLV3	09CLV3 pasteurella
34	7	0.8	88	16	08Z102	08Z102 yersinia pe
35	7	0.8	91	16	09USV0	09USV0 neisseria m
36	7	0.8	95	12	098552	098552 parametecium
37	7	0.8	98	4	09UP45	09UP45 homo sapien
38	7	0.8	105	17	096YV4	096YV4 sulfolobus
39	7	0.8	119	16	092CX1	092CX1 rickettsia
40	7	0.8	119	17	028742	028742 archaeoglob
41	7	0.8	124	13	09PSQ5	09PSQ5 gallus gall
42	7	0.8	126	10	0947P8	0947P8 solanum suc
43	7	0.8	129	16	098NS2	098NS2 rhizobium 1
44	7	0.8	129	16	08UE23	08UE23 agrobacteri
45	7	0.8	138	16	092G5	092G5 rhizobium m
46	7	0.8	130	16	092DB6	092DB6 listeria in
47	7	0.8	130	16	08Y8K1	08Y8K1 listeria mo
48	7	0.8	134	10	08W249	08W249 porphyra te
49	7	0.8	136	16	045067	045067 bacillus su
50	7	0.8	138	16	09K1N0	09K1N0 neisseria m
51	7	0.8	138	16	09JMW3	09JMW3 neisseria m
52	7	0.8	144	17	09YB59	09YB59 aeropyrum p
53	7	0.8	156	10	09FCY2	09FCY2 arabidopsi
54	7	0.8	157	10	09LNM0	09LNM0 arabidopsi
55	7	0.8	157	17	09HJM0	09HJM0 halobacteri
56	7	0.8	161	4	09NM11	09NM11 homo sapien
57	7	0.8	161	4	09BS09	09BS09 homo sapien
58	7	0.8	164	16	09X179	09X179 thermotoga
59	7	0.8	166	17	059317	059317 pyrococcus
60	7	0.8	167	17	08ZV66	08ZV66 pyrobaculum
61	7	0.8	169	17	08TVH6	08TVH6 methanopyru
62	7	0.8	171	3	09HGM2	09HGM2 schizosach
63	7	0.8	174	17	09HSP4	09HSP4 halobacteri
64	7	0.8	176	17	08U3R1	08U3R1 pyrococcus
65	7	0.8	177	10	09M9H3	09M9H3 arabidopsi
66	7	0.8	177	16	0915D8	0915D8 pseudomonas
67	7	0.8	180	16	098G25	098G25 rhizobium 1
68	7	0.8	181	5	025405	025405 lymnaea sta
69	7	0.8	186	16	08XX44	08XX44 ralsomla s
70	7	0.8	191	10	0949H0	0949H0 hordeum vul
71	7	0.8	192	10	09XET4	09XET4 secale cere
72	7	0.8	194	8	09MHV4	09MHV4 ascaris lum
73	7	0.8	194	8	09MHV3	09MHV3 baylisascar
74	7	0.8	194	8	09MHV2	09MHV2 baylisascar
75	7	0.8	194	8	09MHU3	09MHU3 parascaris
76	7	0.8	194	10	09M9G8	09M9G8 arabidopsi
77	7	0.8	198	16	09F3H5	09F3H5 streptomyce
78	7	0.8	199	4	096NC0	096NC0 homo sapien
79	7	0.8	199	11	09CPW7	09CPW7 mus muscu
80	7	0.8	200	13	091RA5	091RA5 gallus gall
81	7	0.8	202	16	09PBG7	09PBG7 xylella fas
82	7	0.8	204	2	086212	086212 enterococu
83	7	0.8	206	10	09LVA5	09LVA5 arabidopsi
84	7	0.8	206	16	097JH2	097JH2 clostridium
85	7	0.8	211	16	08XG83	08XG83 salmone
86	7	0.8	212	16	08RVU8	08RVU8 denrococcus
87	7	0.8	213	16	097T44	097T44 streptococ
88	7	0.8	214	5	062188	062188 caenorhabd1
89	7	0.8	216	17	08TZJ9	08TZJ9 pyrococcus

90	7	0.8	217	16	092M70	092M70 rhizobium m	163	7	0.8	352	16	09KRT9	09KRT9 vibrio chol
91	7	0.8	224	16	09KTL1	09KTL1 neisseria m	164	7	0.8	355	10	09LE25	09LE25 chlamydom
92	7	0.8	225	16	097RT1	097RT1 streptococ	165	7	0.8	355	17	09HS43	09HS43 halobacteri
93	7	0.8	229	17	08ZRT9	08ZRT9 pyrobaculum	166	7	0.8	356	10	064647	064647 arabidopsis
94	7	0.8	230	2	0936W0	0936W0 pseudomonas	167	7	0.8	360	16	09R065	09R065 streptomyc
95	7	0.8	231	16	09HUC7	09HUC7 neisseria m	168	7	0.8	363	2	09EV21	09EV21 bradyrhizob
96	7	0.8	239	16	09JWY3	09JWY3 arabidopsis m	169	7	0.8	365	13	09OZ18	09OZ18 cynops pyr
97	7	0.8	240	10	09LNO2	09LNO2 ralstonia s	170	7	0.8	368	16	09ZES8	09ZES8 listeria in
98	7	0.8	241	16	08YOD6	08YOD6 escherichia	171	7	0.8	368	16	08Y9D6	08Y9D6 listeria mo
99	7	0.8	242	16	08ZV70	08ZV70 salmonella	172	7	0.8	368	16	09RK10	09RK10 streptomyc
100	7	0.8	243	16	08ZPL3	08ZPL3 salmonella	173	7	0.8	370	5	001383	001383 drosophila
101	7	0.8	245	16	08Z6S1	08Z6S1 yersinia pe	174	7	0.8	371	16	09KE08	09KE08 bacillus ha
102	7	0.8	243	16	08Z6S1	08Z6S1 yersinia pe	175	7	0.8	371	16	098N04	098N04 rhizobium l
103	7	0.8	249	17	08ZRT1	08ZRT1 streptomyc	176	7	0.8	371	16	098ZG3	098ZG3 rhizobium l
104	7	0.8	251	2	09XEP6	09XEP6 bacillus me	177	7	0.8	371	16	08XNM2	08XNM2 clostridium
105	7	0.8	251	2	09AMH3	09AMH3 streptomyc	178	7	0.8	372	16	08XNM2	08XNM2 clostridium
106	7	0.8	251	2	09F5J0	09F5J0 streptomyc	179	7	0.8	372	16	09KY31	09KY31 streptomyc
107	7	0.8	251	17	09YFW6	09YFW6 aeropyrum p	180	7	0.8	379	16	09Z724	09Z724 chlamydia p
108	7	0.8	253	2	09X5W5	09X5W5 rhodobacter	181	7	0.8	379	16	09CKD8	09CKD8 pasteurella
109	7	0.8	256	16	08ZS62	08ZS62 anabaena sp	182	7	0.8	382	16	085187	085187 vibrio chol
110	7	0.8	257	16	09KZB8	09KZB8 streptomyc	183	7	0.8	383	2	056725	056725 vibrio algi
111	7	0.8	260	2	093077	093077 rhodococcus	184	7	0.8	383	2	056725	056725 vibrio para
112	7	0.8	266	16	005853	005853 mycobacteri	185	7	0.8	385	16	09T3R2	09T3R2 pseudomonas
113	7	0.8	274	6	08XNL0	08XNL0 clostridium	186	7	0.8	389	16	09AY04	09AY04 caenorhabdi
114	7	0.8	274	6	09CM69	09CM69 oryctolagus	187	7	0.8	389	16	09PGL7	09PGL7 xylella fas
115	7	0.8	277	17	029076	029076 archaeoglob	188	7	0.8	389	16	09CM70	09CM70 pasteurella
116	7	0.8	281	5	08SW14	08SW14 encephalito	189	7	0.8	392	2	09L8Q4	09L8Q4 pseudomonas
117	7	0.8	282	16	09ZTX1	09ZTX1 rhizobium m	190	7	0.8	392	5	08T1A3	08T1A3 drosophila
118	7	0.8	285	17	09Z302	09Z302 archaeoglob	191	7	0.8	394	5	08T1A3	08T1A3 trypanosoma
119	7	0.8	289	10	09ZTX0	09ZTX0 plasm sativ	192	7	0.8	395	16	09BSW2	09BSW2 homo sapien
120	7	0.8	290	11	09CTD9	09CTD9 mus musculu	193	7	0.8	396	16	P72972	P72972 synchocyst
121	7	0.8	291	5	09N5X2	09N5X2 caenorhabdi	194	7	0.8	397	4	043442	043442 bacillus su
122	7	0.8	293	16	083872	083872 treponema p	195	7	0.8	398	5	024466	024466 drosophila
123	7	0.8	295	5	09VTL8	09VTL8 drosophila	196	7	0.8	398	5	09Y000	09Y000 drosophila
124	7	0.8	296	16	09ZMM7	09ZMM7 rhizobium m	197	7	0.8	402	5	045792	045792 caenorhabdi
125	7	0.8	298	10	094B44	094B44 arabidopsis	198	7	0.8	402	16	08Z9J1	08Z9J1 salmonella
126	7	0.8	298	16	09K1U2	09K1U2 chlamydia p	199	7	0.8	402	16	08XC12	08XC12 escherichia
127	7	0.8	304	16	09ZSG3	09ZSG3 rhizobium m	200	7	0.8	402	16	08XBN3	08XBN3 escherichia
128	7	0.8	305	16	08UTB8	08UTB8 agrobacteri	201	7	0.8	404	16	09AK93	09AK93 streptomyc
129	7	0.8	306	16	09CT61	09CT61 pasteurella	202	7	0.8	408	5	08SUV7	08SUV7 streptomyc
130	7	0.8	310	16	09CH39	09CH39 lactococcus	203	7	0.8	408	13	09PTK6	09PTK6 encephalito
131	7	0.8	313	5	08T895	08T895 clona intes	204	7	0.8	412	2	050371	050371 methylophil
132	7	0.8	313	16	09ZAX6	09ZAX6 listeria in	205	7	0.8	421	5	0950B6	0950B6 drosophila
133	7	0.8	314	16	08R6F5	08R6F5 fusobacteri	206	7	0.8	421	16	09EX44	09EX44 streptomyc
134	7	0.8	316	16	09EMP9	09EMP9 streptomyc	207	7	0.8	424	4	096SG3	096SG3 homo sapien
135	7	0.8	317	2	09AQK0	09AQK0 acinetobact	208	7	0.8	427	16	08RSN6	08RSN6 fusobacteri
136	7	0.8	317	16	0985Q7	0985Q7 rhizobium l	209	7	0.8	429	17	09ZTA6	09ZTA6 thermoplasm
137	7	0.8	317	17	08ZEG9	08ZEG9 pyrobaculum	210	7	0.8	430	10	09FT53	09FT53 arabidopsis
138	7	0.8	319	5	09NKJ7	09NKJ7 leishmania	211	7	0.8	432	16	08ZRT9	08ZRT9 ralstonia s
139	7	0.8	319	16	098DY0	098DY0 rhizobium l	212	7	0.8	432	16	08Z8Q7	08Z8Q7 salmonella
140	7	0.8	321	5	09VUD8	09VUD8 drosophila	213	7	0.8	433	5	091725	091725 drosophila
141	7	0.8	322	16	08XIN3	08XIN3 clostridium	214	7	0.8	433	11	054844	054844 mus musculu
142	7	0.8	323	16	08YU50	08YU50 anabaena sp	215	7	0.8	435	16	08XCM2	08XCM2 escherichia
143	7	0.8	323	16	08XFR7	08XFR7 salmonella	216	7	0.8	441	16	093YR3	093YR3 arabidopsis
144	7	0.8	323	17	08X9K0	08X9K0 escherichia	217	7	0.8	441	16	09ZYS9	09ZYS9 rhizobium m
145	7	0.8	323	17	09HKY3	09HKY3 thermoplasm	218	7	0.8	442	5	061670	061670 drosophila
146	7	0.8	325	10	09EVE1	09EVE1 arabidopsis	219	7	0.8	442	5	061671	061671 drosophila
147	7	0.8	330	3	013303	013303 cryptococcu	220	7	0.8	443	10	09FL25	09FL25 arabidopsis
148	7	0.8	334	16	08XTN8	08XTN8 ralstonia s	221	7	0.8	446	16	09SC77	09SC77 streptomyc
149	7	0.8	337	16	08VNX0	08VNX0 anabaena sp	222	7	0.8	446	16	08XT22	08XT22 ralstonia s
150	7	0.8	337	16	08XRS2	08XRS2 ralstonia s	223	7	0.8	449	4	09NX15	09NX15 homo sapien
151	7	0.8	338	3	012004	012004 saccharomye	224	7	0.8	449	4	093934	093934 homo sapien
152	7	0.8	339	5	09VJG9	09VJG9 drosophila	225	7	0.8	450	3	09P6K7	09P6K7 schizosacch
153	7	0.8	341	16	093MW7	093MW7 staphylococ	226	7	0.8	451	10	09CQE8	09CQE8 arabidopsis
154	7	0.8	343	4	08WU91	08WU91 homo sapien	227	7	0.8	452	16	08RGP4	08RGP4 fusobacteri
155	7	0.8	343	5	018014	018014 caenorhabdi	228	7	0.8	453	8	08WJL7	08WJL7 brachionidi
156	7	0.8	343	16	09H2T3	09H2T3 pseudomonas	229	7	0.8	454	16	09XZL6	09XZL6 streptomyc
157	7	0.8	344	16	08UBN3	08UBN3 agrobacteri	230	7	0.8	455	16	08XOK9	08XOK9 ralstonia s
158	7	0.8	345	10	094JL0	094JL0 oryza sativ	231	7	0.8	456	2	09APM5	09APM5 drosophila
159	7	0.8	347	16	08XBE9	08XBE9 escherichia	232	7	0.8	458	4	09H6G4	09H6G4 homo sapien
160	7	0.8	348	9	080301	080301 bacterioph	233	7	0.8	465	4	096R74	096R74 mus musculu
161	7	0.8	351	2	030390	030390 neisseria m	234	7	0.8	465	16	086374	086374 mycobacteri
162	7	0.8	351	10	095YX6	095YX6 nicotiana t	235	7	0.8	471	16	09AA70	09AA70 caulobacter

236	7	0.8	473	10	Q9M2X5	Q9M2X5 arabidopsis	309	7	0.8	643	17	Q97CP6	Q97CP6 thermoplasma
237	7	0.8	477	16	Q9JTN7	Q9JTN7 neisseria m	310	7	0.8	648	16	Q8XYP3	Q8XYP3 ralstonia s
238	7	0.8	479	2	Q9S3Z3	Q9S3Z3 vibrio fisc	311	7	0.8	652	16	Q9XEP3	Q9XEP3 bacillus ha
239	7	0.8	479	16	Q9ZRS3	Q9ZRS3 rhizobium m	312	7	0.8	652	5	Q9VZK2	Q9VZK2 drosophila
240	7	0.8	480	2	Q54034	Q54034 pseudomonas	313	7	0.8	657	4	Q8TF57	Q8TF57 homo sapien
241	7	0.8	480	16	Q9HW85	Q9HW85 pseudomonas	314	7	0.8	658	2	Q9ZHP9	Q9ZHP9 streptomyc
242	7	0.8	485	16	Q9LOA2	Q9LOA2 streptomyc	315	7	0.8	664	5	Q24140	Q24140 drosophila
243	7	0.8	487	16	Q8XZNO	Q8XZNO ralstonia s	316	7	0.8	673	4	Q96SH4	Q96SH4 homo sapien
244	7	0.8	491	2	Q9JH06	Q9JH06 streptomyc	317	7	0.8	676	2	Q9A158	Q9A158 burkholderi
245	7	0.8	493	10	Q8W0R0	Q8W0R0 sorghum bic	318	7	0.8	679	17	Q29619	Q29619 archaeoglob
246	7	0.8	494	17	Q96VC8	Q96VC8 sulfolobus	319	7	0.8	682	16	Q8YHK1	Q8YHK1 bruceella me
247	7	0.8	496	3	Q9URW9	Q9URW9 schizosacch	320	7	0.8	686	10	Q8S562	Q8S562 phaseolus v
248	7	0.8	496	11	Q9Z3M5	Q9Z3M5 mus musculu	321	7	0.8	687	11	Q91XL2	Q91XL2 mus musculu
249	7	0.8	497	16	Q98BV6	Q98BV6 rhizobium l	322	7	0.8	693	3	Q94590	Q94590 schizosacch
250	7	0.8	497	16	Q8RH29	Q8RH29 fusicobacteri	323	7	0.8	695	5	Q9NJB3	Q9NJB3 drosophila
251	7	0.8	499	16	Q8LYL9	Q8LYL9 anabaena sp	324	7	0.8	695	5	Q9NJB2	Q9NJB2 drosophila
252	7	0.8	505	16	Q8UAA5	Q8UAA5 agrobacteri	325	7	0.8	695	5	Q9NCB1	Q9NCB1 drosophila
253	7	0.8	512	2	Q52135	Q52135 escherichia	326	7	0.8	695	5	Q9NCB3	Q9NCB3 drosophila
254	7	0.8	512	2	Q93FL4	Q93FL4 citrobacter	327	7	0.8	695	5	Q9NCB2	Q9NCB2 drosophila
255	7	0.8	512	2	Q9AJ21	Q9AJ21 escherichia	328	7	0.8	695	5	Q9NCB1	Q9NCB1 drosophila
256	7	0.8	512	2	Q47631	Q47631 escherichia	329	7	0.8	695	5	Q9NCB0	Q9NCB0 drosophila
257	7	0.8	512	16	Q85636	Q85636 escherichia	330	7	0.8	695	5	Q9NCA9	Q9NCA9 drosophila
258	7	0.8	513	2	Q93HP2	Q93HP2 streptomyc	331	7	0.8	695	5	Q9NCA6	Q9NCA6 drosophila
259	7	0.8	514	16	Q9ABT2	Q9ABT2 caulobacter	332	7	0.8	695	5	Q9NCA5	Q9NCA5 drosophila
260	7	0.8	516	10	Q9LND9	Q9LND9 arabidopsis	333	7	0.8	695	5	Q9NCA1	Q9NCA1 drosophila
261	7	0.8	517	8	Q68912	Q68912 streptomyc	334	7	0.8	695	5	Q9NCA0	Q9NCA0 drosophila
262	7	0.8	517	16	Q9ABX2	Q9ABX2 rhizobium l	335	7	0.8	695	5	Q9N2R9	Q9N2R9 drosophila
263	7	0.8	518	16	Q8YCS7	Q8YCS7 bruceella me	336	7	0.8	695	5	Q9N2R8	Q9N2R8 drosophila
264	7	0.8	519	16	Q8XWH5	Q8XWH5 ralstonia s	337	7	0.8	695	5	Q9N2R7	Q9N2R7 drosophila
265	7	0.8	519	12	Q57045	Q57045 rat cytoleg	338	7	0.8	695	5	Q9N2R6	Q9N2R6 drosophila
266	7	0.8	523	16	Q33279	Q33279 mycobacteri	339	7	0.8	695	5	Q9N2R5	Q9N2R5 drosophila
267	7	0.8	524	5	Q9XZ28	Q9XZ28 drosophila	340	7	0.8	695	5	Q9N2R4	Q9N2R4 drosophila
268	7	0.8	528	2	Q9Z180	Q9Z180 streptomyc	341	7	0.8	695	5	Q95WB5	Q95WB5 drosophila
269	7	0.8	528	16	Q06828	Q06828 mycobacteri	342	7	0.8	695	5	Q9NCA2	Q9NCA2 drosophila
270	7	0.8	533	17	Q59530	Q59530 pyrococcus	343	7	0.8	695	5	Q9NCA3	Q9NCA3 drosophila
271	7	0.8	534	10	Q9A0U6	Q9A0U6 oryza sativ	344	7	0.8	695	5	Q9NCA4	Q9NCA4 drosophila
272	7	0.8	548	5	Q01384	Q01384 drosophila	345	7	0.8	695	5	Q9NCA7	Q9NCA7 drosophila
273	7	0.8	548	5	Q9VCA2	Q9VCA2 drosophila	346	7	0.8	695	5	Q9NCA8	Q9NCA8 drosophila
274	7	0.8	555	4	Q8WX95	Q8WX95 homo sapien	347	7	0.8	695	5	Q96441	Q96441 drosophila
275	7	0.8	555	16	Q55534	Q55534 synechocyst	348	7	0.8	695	5	Q9U8E8	Q9U8E8 drosophila
276	7	0.8	559	17	Q9Y8U6	Q9Y8U6 aeropyrum p	349	7	0.8	695	5	Q9U8E7	Q9U8E7 drosophila
277	7	0.8	566	10	Q9SK00	Q9SK00 arabidopsis	350	7	0.8	695	5	Q9U8E6	Q9U8E6 drosophila
278	7	0.8	566	16	Q9X1E2	Q9X1E2 thermotoga	351	7	0.8	695	5	Q9U8E5	Q9U8E5 drosophila
279	7	0.8	567	4	Q9H9K7	Q9H9K7 homo sapien	352	7	0.8	695	5	Q9U8E4	Q9U8E4 drosophila
280	7	0.8	567	4	Q96S06	Q96S06 homo sapien	353	7	0.8	695	5	Q9U8E3	Q9U8E3 drosophila
281	7	0.8	567	4	Q95R48	Q95R48 drosophila	354	7	0.8	695	5	Q9U8E2	Q9U8E2 drosophila
282	7	0.8	567	5	Q9VCA3	Q9VCA3 drosophila	355	7	0.8	695	5	Q9U8E1	Q9U8E1 drosophila
283	7	0.8	567	16	Q912D3	Q912D3 pseudomonas	356	7	0.8	695	5	Q9U8D9	Q9U8D9 drosophila
284	7	0.8	567	16	Q8ZJ01	Q8ZJ01 yersinia pe	357	7	0.8	698	17	Q9ZWM4	Q9ZWM4 rhizobium m
285	7	0.8	574	16	Q9A455	Q9A455 caulobacter	358	7	0.8	698	17	Q26996	Q26996 methanobact
286	7	0.8	581	17	Q96XHO	Q96XHO sulfolobus	359	7	0.8	704	3	Q9P8D9	Q9P8D9 candida tro
287	7	0.8	582	12	Q9OH60	Q9OH60 gallid hep	360	7	0.8	706	5	Q9V417	Q9V417 drosophila
288	7	0.8	584	5	Q61113	Q61113 trypanosoma	361	7	0.8	706	17	Q9YAG6	Q9YAG6 aeropyrum p
289	7	0.8	584	17	Q97YQ6	Q97YQ6 sulfolobus	362	7	0.8	709	2	Q93CX3	Q93CX3 lactobacill
290	7	0.8	588	9	Q64304	Q64304 bacterioph	363	7	0.8	712	16	Q8Y2S0	Q8Y2S0 ralstonia s
291	7	0.8	593	2	Q9EY32	Q9EY32 xanthomonas	364	7	0.8	717	4	Q9N0S2	Q9N0S2 homo sapien
292	7	0.8	597	16	Q8U7M3	Q8U7M3 agrobacteri	365	7	0.8	719	2	Q9P150	Q9P150 aquilex pyr
293	7	0.8	613	16	Q9PF38	Q9PF38 xyella fas	366	7	0.8	719	5	Q9U4G1	Q9U4G1 drosophila
294	7	0.8	613	16	Q9KVM0	Q9KVM0 vibrio chol	367	7	0.8	723	10	Q82670	Q82670 cicer ariet
295	7	0.8	614	16	Q97S07	Q97S07 streptococc	368	7	0.8	728	4	Q96MJ7	Q96MJ7 homo sapien
296	7	0.8	617	16	Q9PD27	Q9PD27 xyella fas	369	7	0.8	728	4	Q8TBM1	Q8TBM1 homo sapien
297	7	0.8	623	16	Q8YE13	Q8YE13 bruceella me	370	7	0.8	728	11	Q70263	Q70263 mus musculu
298	7	0.8	627	10	Q49648	Q49648 arabidopsis	371	7	0.8	735	5	Q85XC4	Q85XC4 drosophila
299	7	0.8	629	3	Q9P6U6	Q9P6U6 neuropept	372	7	0.8	741	16	Q8UFT1	Q8UFT1 agrobacteri
300	7	0.8	629	5	Q16801	Q16801 patencitrol	373	7	0.8	744	13	Q91149	Q91149 notoptilham
301	7	0.8	629	5	Q22720	Q22720 caenorhabdi	374	7	0.8	745	12	Q36374	Q36374 alcelaphine
302	7	0.8	630	10	Q9CAN0	Q9CAN0 arabidopsis	375	7	0.8	752	16	Q9PFZ5	Q9PFZ5 xyella fas
303	7	0.8	630	11	Q99J26	Q99J26 mus musculu	376	7	0.8	762	9	Q9ZKK6	Q9ZKK6 bacterioph
304	7	0.8	632	4	Q9BY20	Q9BY20 homo sapien	377	7	0.8	765	16	Q8XN12	Q8XN12 ciostriidum
305	7	0.8	632	4	Q9Y5K6	Q9Y5K6 homo sapien	378	7	0.8	766	2	Q9AER9	Q9AER9 pasteurilla
306	7	0.8	639	4	Q9Y5K5	Q9Y5K5 kluveromyc	379	7	0.8	767	10	Q9SVF1	Q9SVF1 arabidopsis
307	7	0.8	640	2	Q59863	Q59863 streptomyc	380	7	0.8	769	3	Q8X1E4	Q8X1E4 neurospora
308	7	0.8	641	2	Q9XC8	Q9XC8 streptomyc	381	7	0.8	781	16	Q92K65	Q92K65 rhizobium m

382	7	0.8	793	2	085356	085356 butyri/bri	455	7	0.8	1322	10	081490	081490 arabidopsis
383	7	0.8	793	10	092W18	092W18 chara coral	456	7	0.8	1326	5	09W4N4	094490 drosophila
384	7	0.8	795	16	09KZN8	09KZN8 streptomyc	457	7	0.8	1330	6	097961	097961 vulpes vulp
385	7	0.8	802	5	09VJ74	09VJ74 drosophila	458	7	0.8	1335	5	09VE29	09VE29 drosophila
386	7	0.8	802	10	09ZTJ8	09ZTJ8 lycopersico	459	7	0.8	1336	4	014707	014707 homo sapien
387	7	0.8	804	16	09RWM4	09RWM4 delnocooco	460	7	0.8	1373	4	09Y2F5	09Y2F5 homo sapien
388	7	0.8	811	9	09AZ46	09AZ46 colipnaga k	461	7	0.8	1388	4	09Y4M4	09Y4M4 homo sapien
389	7	0.8	812	16	08XS01	08XS01 ralsionia s	462	7	0.8	1416	5	09VY11	09VY11 drosophila
390	7	0.8	815	10	064481	064481 arabidopsis	463	7	0.8	1475	10	09PF22	09PF22 arabidopsis
391	7	0.8	815	13	P87376	P87376 triturus ca	464	7	0.8	1526	5	019611	019611 caenorhabdi
392	7	0.8	820	4	043188	043188 homo sapien	465	7	0.8	1622	5	006550	006550 cryptospori
393	7	0.8	820	4	09BU08	09BU08 homo sapien	466	7	0.8	1643	5	09VJW1	09VJW1 drosophila
394	7	0.8	820	10	09LUN5	088661 rattus norv	467	7	0.8	1653	5	09NKE8	09NKE8 drosophila
395	7	0.8	822	11	088961	09Btr3 homo sapien	468	7	0.8	1702	5	0950K2	0950K2 caenorhabdi
396	7	0.8	825	4	09BTR3	09H2Y6 pseudomonas	469	7	0.8	1808	2	09ENR2	09ENR2 streptomyc
397	7	0.8	830	16	09H2Y6	09H5V8 arabidopsis	470	7	0.8	1886	11	090Y81	090Y81 mus musculu
398	7	0.8	834	10	09FH06	0964Y7 homo sapien	471	7	0.8	1914	13	091008	091008 gallus gall
399	7	0.8	836	4	09H5V8	0964Y7 homo sapien	472	7	0.8	1939	16	082HV5	082HV5 yersinia pe
400	7	0.8	839	10	09LFF3	0911F3 arabidopsis	473	7	0.8	2024	2	09ENR3	09ENR3 streptomyc
401	7	0.8	846	4	075130	075130 homo sapien	474	7	0.8	2157	4	0960C6	0960C6 homo sapien
402	7	0.8	846	16	08ZE91	08ZE91 yersinia pe	475	7	0.8	2157	4	0960C6	0960C6 homo sapien
403	7	0.8	849	15	09YKR7	09YKR7 human immun	476	7	0.8	2197	12	088876	088876 tomato ring
404	7	0.8	850	15	08UTC3	08UTC3 human immun	477	7	0.8	2233	11	08E089	08E089 rattus norv
405	7	0.8	850	16	08YEL8	08YEL8 bruceella me	478	7	0.8	2653	12	09E124	09E124 cercopithec
406	7	0.8	851	15	0994H5	0994H5 human immun	479	7	0.8	2698	10	09XIR5	09XIR5 arabidopsis
407	7	0.8	851	15	070010	070010 human immun	480	7	0.8	3157	16	050378	050378 mycobacteri
408	7	0.8	852	15	0994R5	0994R5 human immun	481	7	0.8	3317	2	09RAH2	09RAH2 nostoc sp.
409	7	0.8	857	15	0994R5	09KX33 bombyx mori	482	7	0.8	3716	16	08XS40	08XS40 mycobacteri
410	7	0.8	861	5	09NKK3	080468 arabidopsis	483	7	0.8	6889	16	08XS40	08XS40 ralsionia s
411	7	0.8	862	10	080468	092V48 arabidopsis	484	7	0.8	7829	5	018559	018559 caenorhabdi
412	7	0.8	862	10	092V48	09DBX9 mycobacteri	485	7	0.8	9507	2	09EW11	09EW11 streptomyc
413	7	0.8	862	16	09CBX9	09DBX9 arabidopsis	486	7	0.8	9510	2	093NX9	093NX9 streptomyc
414	7	0.8	870	10	09PCD5	09JQJ2 alcaiviorax	487	6	0.7	10917	2	093NM6	093NM6 streptomyc
415	7	0.8	872	2	093UQ2	09JPC2 chlamydia m	488	6	0.7	20	15	085636	085636 moloney mur
416	7	0.8	875	16	09PJC2	08XY49 ralsionia s	489	6	0.7	21	11	061917	061917 mus musculu
417	7	0.8	887	16	08XY49	091XX6 mus musculu	490	6	0.7	25	15	071927	071927 human immun
418	7	0.8	912	11	091XX6	09H333 neurospora	491	6	0.7	31	4	090HM9	090HM9 homo sapien
419	7	0.8	928	3	09HEU3	021336 arabidopsis	492	6	0.7	32	4	032493	032493 bacteroides
420	7	0.8	947	10	023136	09WU36 rattus norv	493	6	0.7	33	8	09T2M9	09T2M9 sus scrofa
421	7	0.8	950	11	09WU36	08SBP1 bacterioph	494	6	0.7	33	13	09P872	09P872 rana catesb
422	7	0.8	961	9	08SBP1	096VE8 ustilago ma	495	6	0.7	39	3	096T51	096T51 schizopyll
423	7	0.8	983	9	096VE8	09GR46 tomato ring	496	6	0.7	42	2	048078	048078 mycoplasma
424	7	0.8	985	12	09ORA6	09GR46 tomato ring	497	6	0.7	43	4	090J35	090J35 homo sapien
425	7	0.8	985	12	09ORA7	09GR46 tomato ring	498	6	0.7	44	6	09T7N7	09T7N7 sus scrofa
426	7	0.8	985	12	09ORA8	09GR46 tomato ring	499	6	0.7	45	11	091203	091203 lactobacill
427	7	0.8	985	12	09ORA8	09USU1 schizosacch	500	6	0.7	46	2	08VL20	08VL20 creponema d
428	7	0.8	987	3	09USU1	036025 salmonella	501	6	0.7	47	8	095CL2	095CL2 podolobium
429	7	0.8	992	3	036025	082745 salmoneia	502	6	0.7	47	8	095CL1	095CL1 podolobium
430	7	0.8	993	16	082745	08U7Y7 agrobacteri	503	6	0.7	47	8	095CK9	095CK9 podolobium
431	7	0.8	996	16	08U7Y7	086519 rice yellow	504	6	0.7	47	8	095CK6	095CK6 pultenaea d
432	7	0.8	999	12	086525	086519 rice yellow	505	6	0.7	48	2	09KJ74	09KJ74 scytonema h
433	7	0.8	999	12	086519	08X914 escherichia	506	6	0.7	48	4	095743	095743 homo sapien
434	7	0.8	1005	16	08X914	08XU9 arabidopsis	507	6	0.7	48	4	095743	095743 homo sapien
435	7	0.8	1006	10	08XU9	085064 peanut stun	508	6	0.7	51	1	039226	039226 arabidopsis
436	7	0.8	1006	12	085064	096527 schistosoma	509	6	0.7	51	1	039226	039226 arabidopsis
437	7	0.8	1011	5	096527	091K42 glycine max	510	6	0.7	52	4	09PH93	09PH93 sulfolobus
438	7	0.8	1012	10	091K24	093C4 drosophila	511	6	0.7	52	4	09PH93	09PH93 sulfolobus
439	7	0.8	1028	5	09V3C4	09VTP9 drosophila	512	6	0.7	54	16	09PGV4	09PGV4 xylella fas
440	7	0.8	1033	11	09VTP9	09VTP9 drosophila	513	6	0.7	55	13	092ND7	092ND7 thermus the
441	7	0.8	1033	11	09VTP9	09VTP9 drosophila	514	6	0.7	55	13	092ND7	092ND7 thermus the
442	7	0.8	1050	10	09VTP9	09VTP9 drosophila	515	6	0.7	55	16	08Y4M2	08Y4M2 anabaeia sp
443	7	0.8	1127	5	08T519	091GN9 oryza sativ	516	6	0.7	55	17	097C70	097C70 thermoplas
444	7	0.8	1132	5	002432	08T519 anopheles g	517	6	0.7	57	2	09ZAT1	09ZAT1 synochococ
445	7	0.8	1132	5	09VNA3	09VNA3 drosophila	518	6	0.7	58	2	09ZAT1	09ZAT1 synochococ
446	7	0.8	1138	16	092KB6	092KB6 rhizobium m	519	6	0.7	59	4	096074	096074 homo sapien
447	7	0.8	1157	16	097F62	097F62 clostridium	520	6	0.7	59	5	08WP24	08WP24 plasmodium
448	7	0.8	1187	13	012965	012965 fuqu rubrip	521	6	0.7	60	10	093387	093387 nicotiana t
449	7	0.8	1208	5	09U4W1	09U4W1 aedes aegypt	522	6	0.7	61	2	054538	054538 streptococ
450	7	0.8	1233	3	002979	002979 saccharomyc	523	6	0.7	61	2	093PR1	093PR1 bradyrhizob
451	7	0.8	1236	5	09GRM1	09GRM1 leishmania	524	6	0.7	61	10	09W7X7	09W7X7 arabidopsis
452	7	0.8	1286	10	080725	080725 arabidopsis	525	6	0.7	63	16	09PD64	09PD64 xylella fas
453	7	0.8	1292	10	09M109	09M109 arabidopsis	526	6	0.7	64	16	09PD64	09PD64 xylella fas
454	7	0.8	1300	4	013999	013999 homo sapien	527	6	0.7	66	16	08X3N6	08X3N6 escherichia



528	6	0.7	67	16	08XSY4	08XSY4 ralstonia s	601	6	0.7	89	2	045189	045189 borrelia ga
529	6	0.7	68	0	09Z440	09Z440 pseudomonas	602	6	0.7	89	2	045190	045190 borrelia ga
530	6	0.7	68	6	028692	028692 oryctolagus	603	6	0.7	89	2	045188	045188 borrelia ga
531	6	0.7	68	16	082166	082166 yersinia pe	604	6	0.7	89	16	08REB6	08REB6 fusobacteri
532	6	0.7	69	2	0917G2	0917G2 synechococc	605	6	0.7	89	17	09HPU4	09HPU4 halobacteri
533	6	0.7	69	2	051902	051902 proteus mir	606	6	0.7	89	17	0972M7	0972M7 sulfolobus
534	6	0.7	69	5	09VN37	09VN37 drosophila	607	6	0.7	90	2	032473	032473 pseudomonas
535	6	0.7	70	5	09VRI6	09VRI6 drosophila	608	6	0.7	90	5	P91520	P91520 caenorhabdi
536	6	0.7	70	12	09E4B9	09E4B9 hepatitis b	609	6	0.7	91	11	08R412	08R412 mus musculi
537	6	0.7	70	12	067720	067720 human adeno	610	6	0.7	91	2	056426	056426 thermus the
538	6	0.7	71	2	09L710	09L710 synechococc	611	6	0.7	91	4	014075	014075 homo sapien
539	6	0.7	72	2	09LH33	09LH33 plesiomonas	612	6	0.7	91	5	09VC9	09VC9 drosophila
540	6	0.7	72	12	091S40	091S40 regina rana	613	6	0.7	91	10	048605	048605 hordem vul
541	6	0.7	72	15	083373	083373 murine leux	614	6	0.7	91	15	056583	056583 mlv-like en
542	6	0.7	72	16	092JL4	092JL4 rickettsia	615	6	0.7	91	15	0926G7	0926G7 rickettsia
543	6	0.7	73	2	085925	085925 sphingomona	616	6	0.7	91	16	09RY58	09RY58 delnoccocus
544	6	0.7	73	13	073777	073777 gallus gall	617	6	0.7	92	10	08W217	08W217 setaria fab
545	6	0.7	73	16	08R5T5	08R5T5 thermomaneer	618	6	0.7	92	10	08W214	08W214 setaria ita
546	6	0.7	74	4	09NRX6	09NRX6 homo sapien	619	6	0.7	92	10	08W2H4	08W2H4 setaria ver
547	6	0.7	74	8	037408	037408 emericeia	620	6	0.7	92	16	09K6W5	09K6W5 bacillus ha
548	6	0.7	74	8	037404	037404 allomyces m	621	6	0.7	92	16	053745	053745 mycobacteri
549	6	0.7	74	10	08SAY4	08SAY4 oryza sativ	622	6	0.7	93	2	066021	066021 pseudomonas
550	6	0.7	74	11	09D863	09D863 mus musculi	623	6	0.7	93	8	035964	035964 oryza sativ
551	6	0.7	74	16	09AB64	09AB64 caulobacter	624	6	0.7	93	16	08XEA4	08XEA4 escherichia
552	6	0.7	75	2	08X448	08X448 escherichia	625	6	0.7	93	16	P76067	P76067 escherichia
553	6	0.7	75	2	09AFZ2	09AFZ2 shigella fl	626	6	0.7	94	2	09F2D0	09F2D0 oryza sativ
554	6	0.7	76	2	09JMW3	09JMW3 bradyrhizob	627	6	0.7	94	10	09F7L0	09F7L0 human immun
555	6	0.7	76	10	09SH04	09SH04 arabidopsis	628	6	0.7	94	15	09DYM7	09DYM7 human immun
556	6	0.7	76	16	08XUW0	08XUW0 ralstonia s	629	6	0.7	94	15	09DXY6	09DXY6 human immun
557	6	0.7	76	17	0974E2	0974E2 sulfolobus	630	6	0.7	94	16	P73959	P73959 synechocyst
558	6	0.7	77	5	09VP08	09VP08 drosophila	631	6	0.7	94	17	08RP49	08RP49 methanosarc
559	6	0.7	77	17	08Z211	08Z211 pyrobaculum	632	6	0.7	95	2	007153	007153 mycobacteri
560	6	0.7	78	2	08VSH9	08VSH9 shigella fl	633	6	0.7	95	12	08VAT1	08VAT1 white spot
561	6	0.7	78	16	08VJN9	08VJN9 mycobacteri	634	6	0.7	95	15	003458	003458 feline leux
562	6	0.7	78	17	028207	028207 archaeoglob	635	6	0.7	95	17	08ZUH7	08ZUH7 pyrobaculum
563	6	0.7	79	2	09FPD1	09FPD1 borrelia bu	636	6	0.7	96	2	093705	093705 salmonella
564	6	0.7	79	16	09CC9C	09CC9C mycobacteri	637	6	0.7	96	16	080801	080801 agrobacteri
565	6	0.7	79	16	08ZJK0	08ZJK0 yersinia pe	638	6	0.7	96	17	08R308	08R308 methanosarc
566	6	0.7	79	16	08R968	08R968 thermomaneer	639	6	0.7	97	2	068308	068308 aeromonas s
567	6	0.7	80	12	041946	041946 murid herpe	640	6	0.7	98	10	08S4X8	08S4X8 oryza sativ
568	6	0.7	80	16	08Y3E4	08Y3E4 ralstonia s	641	6	0.7	98	16	09Z8X3	09Z8X3 chlamydia p
569	6	0.7	81	2	085044	085044 thiobacilli	642	6	0.7	98	17	09HPX8	09HPX8 halobacteri
570	6	0.7	81	2	048496	048496 lacobacilli	643	6	0.7	99	10	094ANO	094ANO arabidopsis
571	6	0.7	81	10	024195	024195 oryza sativ	644	6	0.7	100	5	08SRK5	08SRK5 encephalito
572	6	0.7	81	10	024205	024205 oryza sativ	645	6	0.7	100	15	P89820	P89820 human immun
573	6	0.7	81	15	090284	090284 human immun	646	6	0.7	100	15	P89820	P89820 human immun
574	6	0.7	81	15	P90285	P90285 human immun	647	6	0.7	100	17	09Y9S0	09Y9S0 human immun
575	6	0.7	81	16	098FM3	098FM3 rhizobium i	648	6	0.7	101	2	051826	051826 aeropyrum p
576	6	0.7	81	16	08ZGW6	08ZGW6 yersinia pe	649	6	0.7	101	5	09VW77	09VW77 shigella fl
577	6	0.7	82	16	091625	091625 pseudomonas	650	6	0.7	101	5	08SX73	08SX73 drosophila
578	6	0.7	82	16	08YFK4	08YFK4 bruceella me	651	6	0.7	101	12	08UYE7	08UYE7 tt virus. o
579	6	0.7	83	7	09GJ39	09GJ39 damilibac	652	6	0.7	102	2	054771	054771 synechococ
580	6	0.7	83	16	09A931	09A931 caulobacter	653	6	0.7	102	11	064098	064098 mus sp. tru
581	6	0.7	84	17	08THP6	08THP6 methanosarc	654	6	0.7	102	16	09EAV3	09EAV3 rhizobium i
582	6	0.7	84	6	095LE7	095LE7 canis famli	655	6	0.7	103	2	09FDP29	09FDP29 burkholderi
583	6	0.7	85	4	09H513	09H513 homo sapien	656	6	0.7	103	10	004824	004824 tipuana tip
584	6	0.7	85	13	09PWL7	09PWL7 brachydanio	657	6	0.7	103	13	090224	090224 aplocheilic
585	6	0.7	85	16	08RGY2	08RGY2 fusobacteri	658	6	0.7	103	16	08YVK8	08YVK8 ralstonia s
586	6	0.7	86	16	09JY83	09JY83 neisseria m	659	6	0.7	104	2	09AUM8	09AUM8 shigella fl
587	6	0.7	86	16	08UHD5	08UHD5 lacobococcus	660	6	0.7	104	10	09FSQ4	09FSQ4 oryza sativ
588	6	0.7	87	2	048505	048505 lacobococcus	661	6	0.7	104	13	090242	090242 aplocheilic
589	6	0.7	87	3	093NB3	093NB3 lactococcus	662	6	0.7	104	15	091929	091929 human immun
590	6	0.7	87	2	001298	001298 neurospora	663	6	0.7	104	16	099T07	099T07 staphylococ
591	6	0.7	87	15	090AC1	090AC1 human immun	664	6	0.7	104	17	08ZU46	08ZU46 pyrobaculum
592	6	0.7	87	16	08ZLP6	08ZLP6 salmonella	665	6	0.7	105	2	053677	053677 staphylococ
593	6	0.7	87	16	08Z3D9	08Z3D9 salmonella	666	6	0.7	105	2	032926	032926 mycobacteri
594	6	0.7	87	17	08TUZ0	08TUZ0 methanopyru	667	6	0.7	105	6	08WNX4	08WNX4 oryctolagus
595	6	0.7	88	6	046444	046444 dasyurus al	668	6	0.7	105	6	08WNX4	08WNX4 oryctolagus
596	6	0.7	88	10	08W216	08W216 setaria ita	669	6	0.7	105	16	099SM1	099SM1 staphylococ
597	6	0.7	88	16	09KGR2	09KGR2 bacillus ha	670	6	0.7	106	16	09A4D2	09A4D2 caulobacter
598	6	0.7	88	16	0929C0	0929C0 listeria in	671	6	0.7	106	16	08ZED7	08ZED7 yersinia pe
599	6	0.7	88	16	08Y522	08Y522 listeria mo	672	6	0.7	106	16	08RYO0	08RYO0 anabaena sp
600	6	0.7	88	16	08R763	08R763 thermomaneer	673	6	0.7	106	16	09X9V4	09X9V4 streptomyce

674	6	0.7	107	9	Q9FZW8	Q9fzw8 bacterioph	747	6	0.7	119	9	Q9B081	Q9b081 mycobacteri
675	6	0.7	107	16	P71950	P71950 mycobacteri	748	6	0.7	119	16	Q9HXX8	Q9hxx8 pseudomonas
676	6	0.7	107	16	Q9L0H4	Q9l0h4 streptomyc	749	6	0.7	119	16	Q8XP23	Q8xp23 clostridium
677	6	0.7	107	17	Q97Y53	Q97y53 sulfolobus	750	6	0.7	120	16	Q8XP23	Q8xp23 shigella fl
678	6	0.7	108	4	Q9Y533	Q9y533 homo sapien	751	6	0.7	120	2	Q9RPZ7	Q9rpz7 shigella fl
679	6	0.7	108	5	Q9VE99	Q9ve99 drosophila	752	6	0.7	120	5	Q9VU66	Q9vu66 drosophila
680	6	0.7	108	9	Q64354	Q64354 bacterioph	753	6	0.7	120	10	Q8S1U1	Q8s1u1 oryza sativ
681	6	0.7	108	10	Q64354	Q64354 bacterioph	754	6	0.7	120	10	Q8R243	Q8r243 oryza sativ
682	6	0.7	108	10	Q9LXMO	Q9alx0 arabidopsis	755	6	0.7	120	15	Q9ODY5	Q9ody5 porcine end
683	6	0.7	108	16	Q9A316	Q9a316 caulobacter	756	6	0.7	120	16	Q9HJ20	Q9hj20 pseudomonas
684	6	0.7	109	2	Q50470	Q50470 neisseria g	757	6	0.7	120	16	Q8ZKJ9	Q8zkj9 salmonella
685	6	0.7	109	4	Q8WY77	Q8wy77 homo sapien	758	6	0.7	120	16	Q8Z3M7	Q8z3m7 salmonella
686	6	0.7	109	5	Q61467	Q61467 caenorhabdi	759	6	0.7	120	17	Q8Z277	Q8z277 pyrobaculum
687	6	0.7	109	8	Q9T4C7	Q9t4c7 nephroselmi	760	6	0.7	121	8	Q9TLM7	Q9t1m7 cyanidium c
688	6	0.7	109	12	Q9WAY5	Q9way5 tt virus. o	761	6	0.7	121	16	Q9K662	Q9k662 bacillus ha
689	6	0.7	109	16	Q9XE22	Q9xe22 escherichia	762	6	0.7	122	8	Q9MPC4	Q9mfc4 beta vulgar
690	6	0.7	109	16	Q8X4W8	Q8x4w8 escherichia	763	6	0.7	122	10	Q9LH88	Q9lhh8 arabidopsis
691	6	0.7	109	17	Q9YG06	Q9yg06 aeropyrum p	764	6	0.7	122	16	Q9LH88	Q9ldh8 rhizobium l
692	6	0.7	110	1	P95922	P95922 sulfolobus	765	6	0.7	123	3	Q9HDP7	Q9hdp7 collettortic
693	6	0.7	110	5	Q94778	Q94778 trypanosoma	766	6	0.7	123	16	Q92061	Q92g61 rhizobium m
694	6	0.7	110	6	Q46447	Q46447 cercaritelus	767	6	0.7	123	16	Q8YC18	Q8yc18 brucella me
695	6	0.7	110	9	Q8W6M5	Q8w6m5 sinorhizobi	768	6	0.7	124	2	Q68464	Q68464 rickettsia
696	6	0.7	110	10	Q9ARX7	Q9arx7 oryza sativ	769	6	0.7	124	5	Q9U027	Q9u027 giardia lam
697	6	0.7	110	16	Q8XAN0	Q8xan0 escherichia	770	6	0.7	124	16	Q97ON0	Q97gno streptococc
698	6	0.7	111	10	Q42251	Q42251 arabidopsis	771	6	0.7	124	16	Q8ZNB7	Q8znb7 salmonella
699	6	0.7	111	12	Q990T2	Q990t2 hepatitis c	772	6	0.7	124	16	Q8Z4Z8	Q8z4z8 salmonella
700	6	0.7	111	17	Q9YF60	Q9yfe0 aeropyrum p	773	6	0.7	124	17	Q8Z447	Q8z447 archaeoglob
701	6	0.7	112	3	Q9KX29	Q9kx29 pseudomonas	774	6	0.7	125	2	Q68467	Q68467 rickettsia
702	6	0.7	112	3	Q05662	Q05662 saccharomyc	775	6	0.7	125	16	Q8Z350	Q8z350 salmonella
703	6	0.7	112	4	Q9BMU8	Q9bm8 homo sapien	776	6	0.7	125	17	Q97X82	Q97x82 sulfolobus
704	6	0.7	112	4	Q9B216	Q9b216 homo sapien	777	6	0.7	125	2	Q9F9Y3	Q9fy93 neisseria c
705	6	0.7	112	12	Q993S5	Q993s5 banana mild	778	6	0.7	126	2	Q9AMU6	Q9am6 bradyrhizob
706	6	0.7	112	12	Q8UYB9	Q8uyb9 tt virus. o	779	6	0.7	126	5	Q8SS64	Q8sse4 encephalito
707	6	0.7	113	16	Q8YRC7	Q8yrc7 anabaena sp	780	6	0.7	126	16	Q9A8U8	Q9a8u8 caulobacter
708	6	0.7	113	4	Q8BPP3	Q8bpx9 homo sapien	781	6	0.7	126	16	Q9ARK9	Q9ark9 streptomyc
709	6	0.7	113	4	Q9BMU7	Q9bmv0 homo sapien	782	6	0.7	126	16	Q8XT22	Q8xt22 raistonia s
710	6	0.7	113	4	Q9BMU7	Q9bm9 homo sapien	783	6	0.7	127	4	Q96L29	Q96l29 homo sapien
711	6	0.7	113	11	Q81232	Q81232 zea mays (m	784	6	0.7	127	5	Q9NB70	Q9nb70 mesocricetu
712	6	0.7	113	11	Q921U3	Q921u3 mus musculu	785	6	0.7	127	5	Q8RT28	Q8rt28 rickettsia
713	6	0.7	113	16	Q926A0	Q926a0 listeria in	786	6	0.7	127	11	P97282	P97282 mesocricetu
714	6	0.7	113	16	Q92589	Q92589 streptomyc	787	6	0.7	127	16	Q33192	Q33192 mycobacteri
715	6	0.7	114	2	Q87647	Q87647 salmonella	788	6	0.7	128	3	Q9HFP7	Q9hfp7 coccidioid
716	6	0.7	114	2	Q9X5X5	Q9x5x5 mycobacteri	789	6	0.7	128	5	Q16918	Q16918 anopheles a
717	6	0.7	114	2	Q9R9D9	Q9r9d9 bacillus su	790	6	0.7	128	10	Q9LVP6	Q9lvp6 arabidopsis
718	6	0.7	114	3	Q08958	Q08958 saccharomyc	791	6	0.7	128	12	Q8VA19	Q8va19 white spot
719	6	0.7	114	10	Q9AR39	Q9ar39 leavenworth	792	6	0.7	128	15	Q85652	Q85652 murine leuk
720	6	0.7	114	10	Q9AKR5	Q9akr5 leavenworth	793	6	0.7	128	16	Q8ZID9	Q8zid9 rhizobium l
721	6	0.7	114	10	Q9AKR4	Q9akr4 leavenworth	794	6	0.7	129	10	Q9FRI9	Q9fri9 yersinia pe
722	6	0.7	114	10	Q9AXR3	Q9axr3 leavenworth	795	6	0.7	129	16	Q8YHNS	Q8yhn5 arabidopsis
723	6	0.7	114	10	Q9AXR1	Q9axr1 leavenworth	796	6	0.7	129	16	Q52884	Q52884 brucella me
724	6	0.7	114	10	Q9SIC8	Q9sic8 arabidopsis	797	6	0.7	129	16	Q9ZK41	Q9zk41 rhizobium m
725	6	0.7	114	10	Q8S5V7	Q8s5v7 oryza sativ	798	6	0.7	130	12	Q9JFMO	Q9jfm human adeno
726	6	0.7	114	15	Q36580	Q36580 multiple sc	799	6	0.7	130	12	Q9JFMO	Q9jfm human adeno
727	6	0.7	114	16	Q9AAS5	Q9aas5 caulobacter	800	6	0.7	130	17	Q29186	Q29186 archaeglob
728	6	0.7	114	16	Q8XUA5	Q8xus5 raistonia s	801	6	0.7	130	17	Q29186	Q29186 archaeglob
729	6	0.7	115	2	Q9JPE1	Q9jpe1 neisseria m	802	6	0.7	131	2	Q9S387	Q9s387 listeria mo
730	6	0.7	115	2	Q8WTC5	Q8wtc5 helicobacte	803	6	0.7	131	11	Q9DP63	Q9dp63 mus musculu
731	6	0.7	115	2	Q8RM12	Q8rm12 clostridium	804	6	0.7	131	16	Q9RTU3	Q9rtu3 deinococcus
732	6	0.7	115	8	Q98S99	Q98s99 guillardia	805	6	0.7	131	16	Q9AVV1	Q9avv1 caulobacter
733	6	0.7	115	10	Q43762	Q43762 hordenum vul	806	6	0.7	131	16	Q8ZAV3	Q8zav3 yersinia pe
734	6	0.7	115	16	Q43762	Q43762 hordenum vul	807	6	0.7	132	4	Q9UG94	Q9ug94 homo sapien
735	6	0.7	116	10	Q93JPE	Q93jpe ruminococcu	808	6	0.7	132	4	Q9UG94	Q9ug94 homo sapien
736	6	0.7	116	10	Q9AS93	Q9as93 oryza sativ	809	6	0.7	132	5	Q9NPA1	Q9npa1 plasmodium
737	6	0.7	116	12	Q9LFB4	Q9lfb4 chilo lride	810	6	0.7	132	11	Q61638	Q61638 mus musculu
738	6	0.7	116	16	Q9HYO0	Q9hyo0 pseudomonas	811	6	0.7	133	9	Q9U241	Q9u241 caenorhabdi
739	6	0.7	116	16	Q9S1Y6	Q9s1y6 streptomyc	812	6	0.7	133	11	Q9C3M0	Q9c3m0 straphylococ
740	6	0.7	117	12	Q8V7D2	Q8v7d2 tt virus. o	813	6	0.7	133	12	Q9CYO8	Q9cyo8 mus musculu
741	6	0.7	117	12	Q8UYC8	Q8uyc8 tt virus. o	814	6	0.7	133	11	Q9CYO8	Q9cyo8 mus musculu
742	6	0.7	118	2	Q9APX6	Q9apx6 pseudomonas	815	6	0.7	133	16	Q55443	Q55443 ovine parai
743	6	0.7	118	2	Q9FCY9	Q9fcy9 erwinia ste	816	6	0.7	133	16	Q8YMK4	Q8ymk4 anabaena sp
744	6	0.7	118	2	Q46672	Q46672 escherichia	817	6	0.7	133	16	Q8XHQ1	Q8xhq1 clostridium
745	6	0.7	118	6	Q95JY4	Q95jy4 macaca fasc	818	6	0.7	133	16	Q93JY4	Q93jy4 streptomyc
746	6	0.7	118	17	Q8Z248	Q8z248 pyrobaculum	819	6	0.7	133	16	Q98QN4	Q98qn4 mycoplasma

820	6	0.7	133	16	Q92P06	Q92P06 rhizobium m	893	6	0.7	143	12	Q91N36	Q91N36 hepatitis b
821	6	0.7	134	2	Q9K2H4	Q9K2H4 bacillus li	894	6	0.7	143	16	Q93R87	Q93R87 salmonella
822	6	0.7	134	2	Q9K2H3	Q9K2H3 bacillus so	895	6	0.7	143	16	Q8XAT8	Q8XAT8 escherichia
823	6	0.7	134	2	Q9L788	Q9L788 bacillus so	896	6	0.7	143	17	Q9TFH9	Q9TFH9 aeropyrum p
824	6	0.7	134	2	Q9L787	Q9L787 bacillus so	897	6	0.7	144	10	Q9CAM0	Q9CAM0 arabidopsis
825	6	0.7	134	2	Q9L786	Q9L786 bacillus li	898	6	0.7	145	6	Q9BF81	Q9BF81 didelphis m
826	6	0.7	134	2	Q49963	Q49963 mycobacteri	899	6	0.7	145	6	Q9BF80	Q9BF80 macropus eu
827	6	0.7	134	3	Q960R4	Q960R4 coccidioid	900	6	0.7	145	10	Q85181	Q85181 oryza sativ
828	6	0.7	134	4	Q9NV29	Q9NV29 homo sapien	901	6	0.7	145	17	Q27382	Q27382 methanobact
829	6	0.7	134	4	Q96FZ0	Q96FZ0 homo sapien	902	6	0.7	146	2	Q9EVZ8	Q9EVZ8 streptococc
830	6	0.7	134	9	Q03934	Q03934 bacterioph	903	6	0.7	146	5	Q9BPL6	Q9BPL6 metagonimus
831	6	0.7	134	16	Q8UGT2	Q8UGT2 agrobacteri	904	6	0.7	146	9	Q9B0B9	Q9B0B9 mycobacteri
832	6	0.7	134	17	Q8ZMQ0	Q8ZMQ0 pyrobaculum	905	6	0.7	146	10	Q9LPA9	Q9LPA9 arabidopsis
833	6	0.7	135	2	Q9X703	Q9X703 corynebacte	906	6	0.7	146	16	Q9JXV2	Q9JXV2 neisseria m
834	6	0.7	135	16	Q9A908	Q9A908 caulobacte	907	6	0.7	146	16	Q9JWL8	Q9JWL8 neisseria m
835	6	0.7	135	16	Q9XA32	Q9XA32 streptomyce	908	6	0.7	146	16	Q9H235	Q9H235 pseudomonas
836	6	0.7	135	17	Q8U0E7	Q8U0E7 pyrococcus	909	6	0.7	146	16	Q8H046	Q8H046 streptomyce
837	6	0.7	136	5	Q9NMF9	Q9NMF9 leishmania	910	6	0.7	147	2	Q85781	Q85781 myxococcus
838	6	0.7	136	10	Q41859	Q41859 zea mays (m	911	6	0.7	147	4	Q9H657	Q9H657 homo sapien
839	6	0.7	136	11	Q9D8S0	Q9D8S0 mus musculu	912	6	0.7	147	11	Q9CRM2	Q9CRM2 mus musculu
840	6	0.7	136	12	Q8QVW3	Q8QVW3 sen virus.	913	6	0.7	147	16	Q8BDEC	Q8BDEC agrobacteri
841	6	0.7	136	12	Q8QVW2	Q8QVW2 sen virus.	914	6	0.7	148	2	Q9S3M2	Q9S3M2 bordetella
842	6	0.7	136	12	Q8QVW1	Q8QVW1 sen virus.	915	6	0.7	148	2	Q45805	Q45805 clostridium
843	6	0.7	137	2	Q54029	Q54029 propionigen	916	6	0.7	148	2	Q52867	Q52867 streptidum
844	6	0.7	137	2	Q86989	Q86989 proteus mir	917	6	0.7	148	9	Q83345	Q83345 lactococcus
845	6	0.7	137	16	Q9R2T7	Q9R2T7 delnoccocus	918	6	0.7	148	10	Q9SKH0	Q9SKH0 arabidopsis
846	6	0.7	138	3	Q9AJ13	Q9AJ13 escherichia	919	6	0.7	148	10	Q9FTL3	Q9FTL3 arabidopsis
847	6	0.7	138	3	Q9HP60	Q9HP60 ashbya goss	920	6	0.7	148	12	Q91BS1	Q91BS1 turkey heip
848	6	0.7	138	10	Q93ZC6	Q93ZC6 arabidopsis	921	6	0.7	148	16	Q9CLS6	Q9CLS6 pasteurilla
849	6	0.7	138	12	Q8ON95	Q8ON95 ectocarpus	922	6	0.7	148	16	Q970S3	Q970S3 streptococc
850	6	0.7	138	13	Q93606	Q93606 oncorhynch	923	6	0.7	148	17	Q97Y63	Q97Y63 sulfolobus
851	6	0.7	138	16	Q8X603	Q8X603 escherichia	924	6	0.7	149	5	Q96601	Q96601 caenorhabdi
852	6	0.7	138	17	Q97W29	Q97W29 sulfolobus	925	6	0.7	149	10	Q9M209	Q9M209 arabidopsis
853	6	0.7	139	8	Q94RU8	Q94RU8 trypanosoma	926	6	0.7	149	16	Q8YR04	Q8YR04 anabena sp
854	6	0.7	139	8	Q94RU6	Q94RU6 microplitis	927	6	0.7	149	16	Q9KXK8	Q9KXK8 streptomyce
855	6	0.7	139	9	Q64352	Q64352 bacterioph	928	6	0.7	150	2	Q93PM5	Q93PM5 streptococc
856	6	0.7	139	9	Q92X64	Q92X64 mycobacteri	929	6	0.7	150	16	Q25203	Q25203 helicobacte
857	6	0.7	139	9	Q06913	Q06913 brassica na	930	6	0.7	150	16	Q9PM45	Q9PM45 campylobact
858	6	0.7	139	10	Q8VZx3	Q8VZx3 chlamydomon	931	6	0.7	150	16	Q9A7J5	Q9A7J5 caulobacter
859	6	0.7	139	10	Q8VZx3	Q8VZx3 caulobacter	932	6	0.7	151	2	Q47630	Q47630 escherichia
860	6	0.7	139	16	Q9A8A7	Q9A8A7 mus musculu	933	6	0.7	151	2	Q91A89	Q91A89 aeromonas h
861	6	0.7	139	16	Q92JY2	Q92JY2 rhizobium m	934	6	0.7	151	2	Q93FL3	Q93FL3 citrobacter
862	6	0.7	139	17	Q8TKU2	Q8TKU2 methanosarc	935	6	0.7	151	4	Q9H6C7	Q9H6C7 homo sapien
863	6	0.7	140	2	Q9FIW3	Q9FIW3 streptomyce	936	6	0.7	151	5	Q85YP4	Q85YP4 dtrosophila
864	6	0.7	140	2	Q9E0L8	Q9E0L8 corynebacte	937	6	0.7	152	16	Q9KRU5	Q9KRU5 vibrio chol
865	6	0.7	140	2	Q9FA65	Q9FA65 azoarcus ev	938	6	0.7	152	16	Q91S25	Q91S25 pseudomonas
866	6	0.7	140	2	Q9EY95	Q9EY95 corynebacte	939	6	0.7	153	2	Q9F7S1	Q9F7S1 uncultured
867	6	0.7	140	2	Q8VOY9	Q8VOY9 myxococcus	940	6	0.7	153	16	Q9PAN6	Q9PAN6 xylella fas
868	6	0.7	140	4	Q9P0F6	Q9P0F6 homo sapien	941	6	0.7	153	16	Q9JY03	Q9JY03 neisseria m
869	6	0.7	140	5	P91191	P91191 caenorhabdi	942	6	0.7	153	16	Q9JTP4	Q9JTP4 neisseria m
870	6	0.7	140	11	Q9D7H1	Q9D7H1 mus musculu	943	6	0.7	153	16	Q931I0	Q931I0 staphylococ
871	6	0.7	140	16	Q8VJED	Q8VJED pyrococcus	944	6	0.7	153	16	Q8RFM7	Q8RFM7 fusbacteri
872	6	0.7	140	17	Q59212	Q59212 pyrococcus	945	6	0.7	153	17	Q9HRP5	Q9HRP5 halobacteri
873	6	0.7	141	2	Q47941	Q47941 streptococc	946	6	0.7	153	17	Q9HKT6	Q9HKT6 thermoplasma
874	6	0.7	141	9	Q9M8K7	Q9M8K7 bacterioph	947	6	0.7	153	17	Q8TUD5	Q8TUD5 methanosarc
875	6	0.7	141	12	Q72162	Q72162 tobacco rat	948	6	0.7	154	5	Q3VM41	Q3VM41 dtrosophila
876	6	0.7	141	16	Q9RVS8	Q9RVS8 delnoccocus	949	6	0.7	154	8	Q32248	Q32248 gigartina a
877	6	0.7	141	16	Q8XLV9	Q8XLV9 clostridium	950	6	0.7	154	10	Q82011	Q82011 lycopersico
878	6	0.7	142	2	Q66363	Q66363 unidentified	951	6	0.7	154	10	Q82012	Q82012 lycopersico
879	6	0.7	142	2	Q70079	Q70079 unidentified	952	6	0.7	154	12	Q99A06	Q99A06 tt virus.
880	6	0.7	142	10	Q91LV71	Q91LV71 arabidopsis	953	6	0.7	154	12	Q99A06	Q99A06 tt virus.
881	6	0.7	142	10	Q9FHC3	Q9FHC3 arabidopsis	954	6	0.7	154	16	Q9H7M6	Q9H7M6 pseudomonas
882	6	0.7	142	12	Q9DKP6	Q9DKP6 hepatitis b	955	6	0.7	154	16	Q8ZKE6	Q8ZKE6 salmonella
883	6	0.7	142	16	Q8Z349	Q8Z349 salmonella	956	6	0.7	154	16	Q8Z1P8	Q8Z1P8 anabena sp
884	6	0.7	142	16	Q9K3K7	Q9K3K7 streptomyce	957	6	0.7	154	16	Q8YXP9	Q8YXP9 anabena sp
885	6	0.7	142	17	Q8U2E2	Q8U2E2 pyrococcus	958	6	0.7	154	16	Q9X034	Q9X034 streptomyce
886	6	0.7	143	4	Q8WVW5	Q8WVW5 homo sapien	959	6	0.7	154	17	Q9HJK2	Q9HJK2 thermoplasma
887	6	0.7	143	5	Q931I1	Q931I1 anopneles g	960	6	0.7	154	17	Q97WA0	Q97WA0 sulfolobus
888	6	0.7	143	5	Q20125	Q20125 caenorhabdi	961	6	0.7	155	5	Q9V262	Q9V262 aplysia cal
889	6	0.7	143	9	Q8SCP6	Q8SCP6 pseudomonas	962	6	0.7	155	5	Q9V5V9	Q9V5V9 dtrosophila
890	6	0.7	143	10	Q9SPH2	Q9SPH2 gossypium h	963	6	0.7	155	11	Q9CRN0	Q9CRN0 mus musculu
891	6	0.7	143	12	Q91N38	Q91N38 hepatitis b	964	6	0.7	155	16	Q9KXK7	Q9KXK7 streptomyce
892	6	0.7	143	12	Q91N37	Q91N37 hepatitis b	965	6	0.7	156	2	P96135	P96135 thermus the

```

966 6 0.7 156 2 0930W8 093qW8 pseudomonas
967 6 0.7 156 5 0227S7 0227S7 caenorhabdi
968 6 0.7 156 5 09VZNO 09VZNO drosophila
969 6 0.7 156 9 08W6L2 08W6L2 sinorhizobi
970 6 0.7 156 10 08RMG7 08RMG7 arabidopsis
971 6 0.7 156 11 08VDC0 08VDC0 mus musculu
972 6 0.7 156 12 09WTR8 09WTR8 lt virus. o
973 6 0.7 156 16 09JRS9 09JRS9 neisseria m
974 6 0.7 156 17 09YBZ9 09YBZ9 aeropyrum p
975 6 0.7 157 2 0455I2 0455I2 bacillus su
976 6 0.7 157 2 093SL6 093SL6 streptococc
977 6 0.7 157 10 09LSI3 09LSI3 arabidopsis
978 6 0.7 157 16 08XRT5 08XRT5 ralsonia s
979 6 0.7 157 17 09H0I0 09H0I0 halobacteri
980 6 0.7 158 2 09F901 09F901 comamonas s
981 6 0.7 158 10 09M5F1 09M5F1 medicago sa
982 6 0.7 158 10 09C7P9 09C7P9 arabidopsis
983 6 0.7 158 16 09A0F1 09A0F1 streptococc
984 6 0.7 158 16 097PK5 097PK5 streptococc
985 6 0.7 159 2 049778 049778 mycobacteri
986 6 0.7 159 2 08VM06 08VM06 streptococc
987 6 0.7 159 10 095XZ9 095XZ9 arabidopsis
988 6 0.7 159 16 09RUC5 09RUC5 deinococcus
989 6 0.7 159 16 09KSH8 09KSH8 vibrio chol
990 6 0.7 159 16 09A1G2 09A1G2 streptococc
991 6 0.7 159 16 08YPC8 08YPC8 anabaena sp
992 6 0.7 160 2 030595 030595 escherichia
993 6 0.7 160 2 093M08 093M08 xanthomonas
994 6 0.7 160 3 099139 099139 ustilago sc
995 6 0.7 160 3 078566 078566 agaricus bl
996 6 0.7 160 16 031590 031590 bacillus su
997 6 0.7 161 2 030617 030617 pseudomonas
998 6 0.7 161 2 093AN7 093AN7 schistosom
999 6 0.7 161 3 014318 014318 schistosom
1000 6 0.7 161 12 09E997 09E997 hepatitis b

```

## ALIGNMENTS

```

RESULT 1
082B95 PRELIMINARY: PRT: 578 AA.
AC 082B95:
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative exported protein.
GN YP03524.
OS Versinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Versinia.
OX NCBI_TaxID=632:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS.
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdono-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RA "Genome sequence of Versinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL: AJ414157; CAC92753.1; -
DR InterPro: IPR000566; Lipoclin_cytfabp.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SEQUENCE 578 AA: 64405 MW; 69AD3780BDC20F1 CRC64;
Query Match 1.4%: Score 13; DB 16; Length 578;

```

```

Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 880 GAGVGRWASPVG 892
Db 538 GAGVGRWASPVG 550

```

## RESULT 2

```

08XEL5 PRELIMINARY: PRT: 577 AA.
AC 08XEL5:
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative outer membrane protein (Putative exported protein).
GN YPFW OR S7M4409 OR S7Y4768.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Speleth J., Clifton S.W., Latreille P.,
RA Courtney L., Portwillik S., Ali J., Danie M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flora L., Miller W., Stoneking T., Nhan N.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahina M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsels K.,
RA Krogh A., Larsen T.S., Leather S., Hien T.T., Holroyd S., Jagsels K.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL: AE008906; AL23229.1; -
DR InterPro: IPR00184; CAD06889.1; -
DR InterPro: IPR000566; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surfAg; 1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SEQUENCE 577 AA: 64765 MW; FBA4190173BP18C CRC64;
Query Match 1.2%: Score 11; DB 16; Length 577;
Best Local Similarity 100.0%; Pred. No. 0.077;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

## RESULT 3

```

Q91001 PRELIMINARY: PRT: 579 AA.
AC Q91001:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

```

```

DT 01-OCT-2001 (TReMBLrel. 18, last annotation update)
DE Hypothetical protein PA2543.
GN PA2543.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stoyer C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Yim Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT *Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.*
RL Nature 406:959-964(2000).
DR EMBL; AE004682; AAG05931.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 579 AA; 63739 MW; 23467005C836FBF CRC64;

Query Match
Best Local Similarity 1.0%; Score 10; DB 16; Length 579;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 809 LRFAGGDOS 818
DB 470 LRFAGGDOS 479
|||||

RESULT 4
ID 084951 PRELIMINARY; PRT; 260 AA.
AC 084951.
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
DE Ssef (secretion system effector).
GN Ssef OR STM1404.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN
RP SEQUENCE OF 253-260 FROM N.A.
RC STRAIN-SL1344.
RX MEDLINE=97451029; PubMed=9302299;
RA Valdivia R.H., Falkow S.;
RT *Fluorescence-based isolation of bacterial genes expressed within host
RT cells.*;
RL Science 277:2007-2011(1997).
RN
RP SEQUENCE FROM N.A.
RC STRAIN-SL1344.
RX MEDLINE=99000133; PubMed=9786194;
RA Cifillio D.M., Valdivia R.H., Monack D.M., Falkow S.;
RT *Macrophage-dependent induction of the Salmonella pathogenicity island
RT 2 type III secretion system and its role in intracellular survival.*;
RL Mol. Microbiol. 30:175-188(1998).
RN
RP SEQUENCE FROM N.A.
RC STRAIN-SL1344.
RX Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RC STRAIN-LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Lattelle P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

```

```

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT *Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.*;
RL Nature 413:852-856(2001).
DR EMBL; AF020808; AAC28885.1; -.
DR EMBL; AE008761; AAL20328.1; -.
KW Complete proteome.
SQ SEQUENCE 260 AA; 26725 MW; 0890B23F4FC0DE78 CRC64;

Query Match
Best Local Similarity 1.0%; Score 9; DB 16; Length 260;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 PVLTPEOI 66
DB 43 PVLTPEOI 51
|||||

RESULT 5
ID 09R802 PRELIMINARY; PRT; 262 AA.
AC 09R802.
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, last annotation update)
DE Ssef.
GN Ssef.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-LT2.
RX MEDLINE=99000132; PubMed=9786193;
RA Hensel M., Shea J.E., Waterman R., Mundy R., Nikolaus T., Banks G.,
RA Vazquez-Torres A., Gleason C., Fang F.C., Holden D.W.;
RT *Genes encoding putative effector proteins of the type III secretion
RT system of Salmonella pathogenicity island 2 are required for bacterial
RT virulence and proliferation in macrophages.*;
RL Mol. Microbiol. 30:163-174(1998).
DR EMBL; AJ224892; CAA12191.1; -.
SQ SEQUENCE 262 AA; 26893 MW; 62A27ED053DDCEFA CRC64;

Query Match
Best Local Similarity 1.0%; Score 9; DB 2; Length 262;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 58 PVLTPEOI 66
DB 43 PVLTPEOI 51
|||||

RESULT 6
ID 09CK26 PRELIMINARY; PRT; 586 AA.
AC 09CK26.
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE Hypothetical protein PM1809.
GN PM1809.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-PM70.
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;

```

```

RT      "Complete genomic sequence of Pasteurella multocida pm70."
RL      Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR      EMBL: AE006218; AK03893.1; -.
DR      InterPro: IPR0010184; Bac_surfAg_D15.
DR      InterPro: IPR001092; HLH_basic.
DR      Pfam: PF01103; Bac_surfAg_1.
DR      ProSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 586 AA; 67106 MW; 00895174DED58283 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 16; Length 586;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      884 GVRMASPGV 892
Db      549 GVRMASPGV 557

RESULT 7
O9RPH7 PRELIMINARY; PRT; 1085 AA.
AC      O9RPH7;
DT      01-MAY-2000 (TREMBlrel. 13, Created)
DT      01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT      01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE      Recd.
DE      RECC.
GN      Mycobacterium smegmatis.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1772;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=99412429; PubMed=10481025;
RX      Griffin IV T.J., Parsons L., Leshchiner A.E., Devost J.,
RA      Derbyshire K.M., Grindley N.D.;
RT      "In vitro transposition of tm552: a tool for DNA sequencing and
RT      mutagenesis.";
RL      Nucleic Acids Res. 27:3859-3865(1999).
DR      EMBL: AF157643; A046807.1; -.
SQ      SEQUENCE 1085 AA; 120352 MW; F4D9A81AA6A277C4 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 2; Length 1085;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      236 QTAIVARA 244
Db      753 QTAIVARA 761

RESULT 8
P95086 PRELIMINARY; PRT; 174 AA.
AC      P95086;
DT      01-MAY-1997 (TREMBlrel. 03, Created)
DT      01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT      01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE      Hypothetical protein RV3072C.
DE      RV3072C OR MT3157 OR MTCY22D7.09.
OS      Mycobacterium tuberculosis.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1773;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=H37RV;
RX      MEDLINE=98295987; PubMed=9634230;
RA      Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA      Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA      Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

```

```

RA      Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA      Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA      Oliver S., Osborne J., Quail M.A., Rajandram M.A., Rogers J.,
RA      Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA      Sutton J.E., Taylor K., Whitehead S., Barrell B.G.;
RT      "Deciphering the biology of Mycobacterium tuberculosis from the
RT      complete genome sequence.";
RL      Nature 393:537-544(1998).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CDC 1551 / OSHKOSH;
RA      Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA      Peterson J.F., Deboy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA      Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA      Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA      Bishai W.;
RT      "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT      laboratory strains.";
RL      Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL: Z83866; CAB06251.1; -.
DR      EMBL: AE007133; AAK47493.1; ALT_INT.
DR      TIGR: MT3157; -.
DR      TubercuList: RV3072C; -.
DR      InterPro: IPR002103; Bac_Luciferase.
DR      Pfam: PF00296; bac_Luciferase; 1.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 174 AA; 17844 MW; 5073A9E87C0359D CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 174;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      895 RVDVATGV 902
Db      90 RVDVATGV 97

RESULT 9
O98JY3 PRELIMINARY; PRT; 235 AA.
AC      O98JY3;
DT      01-OCT-2001 (TREMBlrel. 18, Created)
DT      01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT      01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE      Hypothetical protein mlr1763.
GN      MLR1763.
OS      Rhizobium loti (Mesorhizobium loti).
OC      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC      Rhizobiaceae; Mesorhizobium.
OX      NCBI_TaxID=381;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=MAFF303099;
RX      MEDLINE=21082930; PubMed=11214968;
RA      Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA      Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA      Kishida Y., Kiyokawa C., Kohara M., Matsuno M., Matsumoto A.,
RA      Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA      Takeuchi C., Yamada M., Tabata S.;
RT      "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT      Mesorhizobium loti.";
RL      DNA Res. 7:331-338(2000).
DR      EMBL: AP002998; BAB49062.1; -.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 235 AA; 26023 MW; DB2C4A2DA6D19FC6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 235;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      238 ALVARAV 245
Db      224 ALVARAV 231

```

```

RESULT 10
Q98HK9 PRELIMINARY: PRT; 319 AA.
ID 098HK9:
AC 098HK9:
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Probable oxidoreductase.
GN ML2825.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shlimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RT DNA Res. 7:331-338(2000).
DR EMBL; AP003000; BAB49857.1; -
DR InterPro: IPR000683; GFO_IDH_MOCA.
DR InterPro: IPR004104; GFO_IDH_MOCA.C.
DR InterPro: IPR002016; Peroxidase.
DR Pfam; PF01408; GFO_IDH_MOCA.1.
DR Pfam; PF02894; GFO_IDH_MOCA.C; 1.
DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 319 AA; 35433 MW; A4F8C2E20977E352 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 319;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 VEVYRGECA 287
Db 181 VEVYRGECA 188

RESULT 11
Q927W8 PRELIMINARY: PRT; 331 AA.
ID 0927W8:
AC 0927W8:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Mbl protein.
GN MBL OR LIN2669.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Chabdit A., Chetouani F., Couve E., de Darvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurupkat G.,
RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,

```

```

RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596173; CAC97895.1; -
DR Listlist: LIN02669; -
DR InterPro: IPR001023; Hsp70.
DR InterPro: IPR004753; MreB_MrL.
DR PRINTS; PR01652; SHAPEPROTEIN.
DR Prodom; PD000089; Hsp70; 1.
DR TIGRfams; TIGR00904; mreB; 1.
KW Complete proteome.
SQ SEQUENCE 331 AA; 35581 MW; 1FB5D97D188CE7EA CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 331;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 839 GOVLAVGT 846
Db 38 GOVLAVGT 45

RESULT 12
Q8Y4C5 PRELIMINARY: PRT; 331 AA.
ID 08Y4C5:
AC 08Y4C5:
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Mbl protein.
GN MBL OR LMO2525.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Chabdit A., Chetouani F., Couve E., de Darvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurupkat G.,
RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591983; CAD00603.1; -
DR Listlist: LMO02525; -
DR InterPro: IPR001023; Hsp70.
DR InterPro: IPR004753; MreB_MrL.
DR PRINTS; PR01652; SHAPEPROTEIN.
DR Prodom; PD000089; Hsp70; 1.
DR TIGRfams; TIGR00904; mreB; 1.
KW Complete proteome.
SQ SEQUENCE 331 AA; 35636 MW; 46A0FFC9CDE13825 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 331;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 839 GOVLAVGT 846
Db 38 GOVLAVGT 45

RESULT 13
Q92BL4

```

ID Q9ZBL4 PRELIMINARY; PRT; 340 AA.  
 AC Q9ZBL4;  
 DT 01-MAY-1999 (TReMBLrel. 10, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Hypothetical 36.3 kDa protein (putative oxidoreductase).  
 GN M0299 OR M0299.24  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies K.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus.";  
 RL Nature 409:1007-1011(2001).  
 DR EMBL: AL035159; CAA22708.1; -;  
 DR EMBL: AL583918; CAC29807.1; -;  
 DR Leproma: M0299; -;  
 DR InterPro: IPR000927; D\_aa-oxidase.  
 DR Pfam: PF01266; DAO; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 340 AA; 36311 MW; E8D64FB7F817658B CRC64;

Query Match  
 Best Local Similarity 0.9%; Score 8; DB 16; Length 340;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 884 GVRMASPV 891  
 |||||||  
 Db 167 GVRMASPV 174

RESULT 14  
 ID P72239 PRELIMINARY; PRT; 379 AA.  
 AC P72239;  
 DT 01-FEB-1997 (TReMBLrel. 02, Created)  
 DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE CFA-beta-ketolactamase.  
 GN CFA3.  
 OS Pseudomonas syringae (pv. glyciniae).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 NX NCBI\_TaxID=318;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GLYCINA;  
 RX MEDLINE=97149295; PubMed=8996103;  
 RA Penfold C.N., Bender C.L., Turner J.G.;  
 RT "Characterisation of genes involved in biosynthesis of coronafacic  
 RT acid, the polyketide component of the phytotoxin coronatine.";  
 RL Gene 183:167-173(1996).  
 DR EMBL: U56980; AAB41300.1; -;  
 DR HSSP: P73283; 1E5M.  
 DR InterPro: IPR000794; ketoacyl-synt.  
 DR Pfam: PF02801; ketoacyl-synt. 1.  
 SQ SEQUENCE 379 AA; 39691 MW; 629745BCE98DA0B1 CRC64;

Query Match 0.9%; Score 8; DB 2; Length 379;  
 Best Local Similarity 100.0%; Pred. No. 60;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 529 DAVSAVAR 536  
 |||||||  
 Db 229 DAVSAVAR 236

RESULT 15  
 ID Q9KYD1 PRELIMINARY; PRT; 383 AA.  
 AC Q9KYD1;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Putative membrane protein.  
 GN SC06899 OR SC182.05.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 NX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Saunders D.C., Harris D.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajadream M.A.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphry L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajadream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL: AL356812; CAB92561.1; -;  
 DR InterPro: IPR000005; HTHARAC.  
 DR PROSITE: PS00041; HTH\_ARAC\_FAMILY 1; UNKNOWN\_1  
 SQ SEQUENCE 383 AA; 37020 MW; 9FFEPD7E5827D4E CRC64;

Query Match 0.9%; Score 8; DB 16; Length 383;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 721 ALLGAVAV 728  
 |||||||  
 Db 170 ALLGAVAV 177

RESULT 16  
 ID Q8V5A8 PRELIMINARY; PRT; 452 AA.  
 AC Q8V5A8;  
 DT 01-MAR-2002 (TReMBLrel. 20, Created)



```

DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Nonstructural protein 2.
GN NS2.
OS Casphalia extranea densovirus.
OC Viruses: ssRNA viruses; Parvoviridae; Densovirinae; Densovirus.
ON NCBI_TaxID=180586;
RN [1]
RP SEQUENCE FROM N.A.
RA Tjissen P., Fediere G., Li Y., Zadori Z., Szelei J.;
RT "Genome organization of Casphalia extranea densovirus (CedNV), a new
RT iteravirus.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF375296; AAL56545.1; -.
SQ SEQUENCE 452 AA; 52387 MW; B0FBA9129C148C7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 12; Length 452;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 TRPISLEE 122
DB 180 TRPISLEE 187

RESULT 17
ID 086499 PRELIMINARY; PRT; 469 AA.
AC 086499:
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein SC06331.
GN SC06331 OR SC10H5.07.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
ON NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL031232; CAA20279.1; -.
KW Hypothetical protein.
SQ SEQUENCE 469 AA; 52166 MW; A9D619820B16389A CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 469;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 PVALAAYL 21
DB 420 PVALAAYL 427

RESULT 18
ID 076799 PRELIMINARY; PRT; 506 AA.
AC 076799:
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)

```

```

DE Microsomal cytochrome P450.
GN CYP9B3.
OS Drosophila mettleri (Fruit fly).
ON OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7228;
RN [1]
RP SEQUENCE FROM N.A.
RA Danielson P.B., Reeves S.A., Fogleman J.C.;
RT "Isolation of a novel cytochrome P450 (CYP9B3) from Drosophila
RT mettleri.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF083945; AAC33297.1; -.
DR FlyBase; FBgn0025418; DmelCYP9B3.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 506 AA; 5855 MW; 246BC0510F6901D0 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 5; Length 506;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 LSNDLIAT 417
DB 185 LSNDLIAT 192

RESULT 19
ID 087139 PRELIMINARY; PRT; 508 AA.
AC 087139:
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Phycocyanin alpha phycocyanobilin lyase related protein.
GN MA4317.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
ON NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C2A / ATCC 35395 / DSM 2834;
RX MEDLINE-21929760; PubMed-11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Galvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeBrellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE011149; AAM07661.1; -.
KW Lyase; Complete proteome.
SQ SEQUENCE 508 AA; 55620 MW; A07C705CBCA80278 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 17; Length 508;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 LQAVRAL5 411
DB 264 LQAVRAL5 271

```

```

RESULT 20
09KRP31
ID 09KRP31 PRELIMINARY: PRT: 582 AA.
AC 09KRP31:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein VC2548.
GN VC2548.
OS Vibrrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Elsen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Esmolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RT Nature 406:477-483(2000).
RL EMBL: AE004323; AAF95689.1; -.
DR TIGR: VC2548.
DR InterPro: IPR000184; Bac_surfAg_D15.
DR InterPro: IPR000566; Lipocin_cytfABP.
DR Pfam: PF01103; Bac_surface_Ag; 1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 582 AA; 65644 MW; E96946B85F1C426 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 582;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 809 LREFPAGD 816
DB 473 LREFPAGD 480

RESULT 21
08ZSS3
ID 08ZSS3 PRELIMINARY: PRT: 731 AA.
AC 08ZSS3:
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein A117659.
GN A117659.
OS Anabaena sp. (strain PCC 7120).
OC Plasmid PCC7120beta.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kurlitz T., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
RX EMBL: AF003602; BAB7302.1; -.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 731 AA; 83004 MW; DEAB8AA4F02DE54D5 CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 731;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 IKALEDI 219
DB 116 IKALEDI 123

RESULT 22
09HK17
ID 09HK17 PRELIMINARY: PRT: 1047 AA.
AC 09HK17:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Probable carbamoyl-phosphate synthase, large subunit.
GN TA0791.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumelster W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum."
RT Nature 407:508-513(2000).
RL EMBL: AL445065; CAC11922.1; -.
DR HSSP: P00968; LJDB.
DR InterPro: IPR000901; CPsase.
DR InterPro: IPR004362; MGS_like.
DR Pfam: PF00289; CPsase_L_chain; 2.
DR Pfam: PF02786; CPsase_L_D2; 1.
DR Pfam: PF02787; CPsase_L_D3; 1.
DR Pfam: PF02142; MGS; 1.
DR PRINTS: PR00098; CPsase.
DR PROSITE: PS00867; CPsase_2; UNKNOWN_2.
DR Complete proteome.
SQ SEQUENCE 1047 AA; 116851 MW; E062CAD131746864 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 17; Length 1047;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 626 KGKLYAT 633
DB 941 KGKLYAT 948

RESULT 23
09VKH2
ID 09VKH2 PRELIMINARY: PRT: 1373 AA.
AC 09VKH2:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE SALM protein.
GN SALM OR CG6464.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

```

RA Amaxatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazel R.G., Champe M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abill J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borikva D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weissendbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003632; AAF53097.1; -.  
 DR HSSP: P15822; 1B80.  
 DR Flybase: FBgn0004579; salm.  
 DR InterPro: IPR0008822; znf.C2H2.  
 DR Pfam: PF00096; zf-C2H2; 7.  
 DR SMART: SM00355; znf.C2H2; 7.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 7.  
 DR PROSITE: PS00157; ZINC\_FINGER\_C2H2\_2; 7.  
 DR DNA-binding: Metal-binding; Zinc-finger.  
 KW DNA-binding; Metal-binding; Zinc-binding.  
 SQ SEQUENCE 1373 AA; 151014 MW; A6D857870F645ACC CRC64;

Query Match 0.9%; Score 8; DB 5; Length 1373;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 61 LTRPQIOA 68  
 Db 913 LTRPQIOA 920  
 RESULT 24  
 048697 PRELIMINARY; PRT; 1417 AA.  
 AC 048697;  
 DT 01-JUN-1998 (TRMBLrel. 06, Created)  
 DT 01-JUN-1998 (TRMBLrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TRMBLrel. 17, Last annotation update)  
 DE F316.24 protein.  
 GN F316.24.  
 GN Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-CV, COLUMBIA;  
 RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,  
 RA Au M., Ataulo R., Buehler E., Dewar K., Feng J., Kim C., Li Y.,  
 RA Ol J.G., Osborne B.L., Shinn P., Sun H., Toriumi M., Vysotskaya V.S.,  
 RA Yu G., Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC002396; AAC00591.1; -.  
 DR InterPro: IPR003169; GYF.  
 DR Pfam: PF02213; GYF; 1.  
 DR SMART: SM00444; GYF; 1.  
 SQ SEQUENCE 1417 AA; 157750 MW; 5B08FBAC1F48334 CRC64;

Query Match 0.9%; Score 8; DB 10; Length 1417;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 387 RELLEQL 394  
 Db 843 RELLEQL 850

RESULT 25  
 045283 PRELIMINARY; PRT; 33 AA.  
 ID 045283;  
 AC 045283;  
 DT 01-NOV-1996 (TRMBLrel. 01, Created)  
 DT 01-NOV-1996 (TRMBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TRMBLrel. 08, Last annotation update)  
 DE Alpha-amylase precursor (Fragment).  
 GN AMYL.  
 OS Bacillus licheniformis.  
 OC Bacteria; Firmicutes; Bacillus/clostridium group; Bacillales;  
 OC Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1402;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91092499; PubMed=2265757;  
 RA Jorgensen P.L., Hansen C.K., Poulsen G.B., Diderichsen B.;  
 RT "In vivo genetic engineering: homologous recombination as a tool for  
 RT plasmid construction.";  
 RL Gene 96:37-41(1990).  
 DR EMBL: M62637; AAA2232.1; -.  
 DR SIGNAL.  
 FT SIGNAL. 1 29  
 FT CHAIN 30 >33  
 FT NON\_TER 33  
 FT SEQUENCE 33 AA; 3711 MW; 9CA3D49C5F5C36F7 CRC64;

Query Match 0.8%; Score 7; DB 2; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 166 KRLYARL 172  
 Db 5 KRLYARL 11  
 RESULT 26  
 08W6C9 PRELIMINARY; PRT; 62 AA.  
 ID 08W6C9;  
 AC 08W6C9;  
 DT 01-MAR-2002 (TRMBLrel. 20, Created)  
 DT 01-MAR-2002 (TRMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TRMBLrel. 20, Last annotation update)  
 DE Menda.  
 GN ORF62.  
 GN Viruslike phage VSK.  
 OC Viruses; ssDNA viruses; Inoviridae; Inovirus.  
 OX NCBI\_TaxID=181604;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Basu N., Kar S., Ghosh R.K.;  
 RT "Molecular analysis of filamentous phage VSK of *Vibrio cholerae* 0139;

```

RT A possible clue to genetic transmission."
DR Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
SQ SEQUENCE 62 AA; 7221 MW; 56813397E7BCBDF CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 888 ASPVGV 894
    |||||
Db 37 ASPVGV 43

RESULT 27
ID 08W542 PRELIMINARY; PRT; 66 AA.
AC 08W542;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Dehydrin (Fragment).
GN DHN2.
OS Retama raetam.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Retama.
OX NCBI_TaxID=49837;
RN [1]
RP SEQUENCE FROM N.A.
RA Pnueli L., Mitter R.;
RT "Dehydrin in retama raetam.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF439276; AAL32035.1; -
DR Interpro: IPR000167; Dehydrin.
DR Pfam: PF00257; dehydrin.1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 66 AA; 7354 MW; 857C8C3DE22912F6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 150 EVVVPPT 156
    |||||
Db 43 EVVVPPT 49

RESULT 28
ID 09PD36 PRELIMINARY; PRT; 66 AA.
AC 09PD36;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein Xf1543.
GN Xf1543.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9ASC;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Agencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.B.A., Carraro D.M., Carter H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,

```

```

RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsumura A.Y.,
RA Merck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Pelxoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
RA de Souza A.P., Terezzi M.F., Truffi D., Tsai S.M., Tsuboko M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL: AE003984; AAF84352.1; -
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 66 AA; 6871 MW; 8DA1217E63550F41 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 809 LRFVAG 815
    |||||
Db 11 LRFVAG 17

RESULT 29
ID P74009 PRELIMINARY; PRT; 72 AA.
AC P74009;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein srr2201.
GN SSR2201.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nairuo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL: D90911; BAA18080.1; -
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 72 AA; 8400 MW; C66AE11FE431E0BF CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 216 LEDITQE 222
    |||||
Db 54 LEDITQE 60

RESULT 30
ID 08XG35

```

```

ID 08XGJ5 PRELIMINARY; PRT: 82 AA.
AC 08XGJ5;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DE Hydrogenase-2 operon protein (Hydrogenase-2 component protein).
GN HtbO OR STM3143 OR STY3314.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
XX [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA Mclelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Kirogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skellon J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR EMBL: AE008844; AAL22017.1; -
DR EMBL: AL672727; CAD02975.1; -
DR InterPro: IPR001109; HUPF_HyPC.
DR Pfam: PF01455; HUPF_HyPC.1.
DR PRINTS: PR00445; HUPFHYP.
DR ProDom: PD003112; HUPF_HyPC.1.
DR TIGRFAMs: TIGR00074; hupc_hyPC.1.
DR PROSITE: PS01097; HUPF_HyPC.1.
KW Complete proteome.
SQ SEQUENCE 82 AA; 8850 MW; 381D32AB941860E7 CRC64;

Query Match 0.8%; Score 7; DB 16; Length 82;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 839 GOVLAVG 845
Db 7 GOVLAVG 13

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932228;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArelano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grigoriev D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnet H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL: AE010791; AAM04658.1; -
KW Complete proteome.
SQ SEQUENCE 84 AA; 8963 MW; 55C35F09C61143C3 CRC64;

Query Match 0.8%; Score 7; DB 17; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 MNVTEIV 428
Db 35 MNVTEIV 41

RESULT 32
ID 09PV18 PRELIMINARY; PRT: 85 AA.
AC 09PV18;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Regulator of G-protein signalling 12 (Frigment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Schiff M., Jordan J.D., Diverse-Pierluisi M.;
RT "Mechanisms of desensitization of G-activated tyrosine kinase
RT pathways."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC 1- SIMILARITY: CONTAINS 1 RGS DOMAIN.
DR EMBL: AF090086; AAF00029.1; -
DR HSSP: P49799; IAGR.
DR InterPro: IPR000342; Regl_Gprotein.
DR Pfam: PF00615; RGS.1.
DR PRINTS: PR01301; RGSPTROTEIN.
DR ProDom: PD001580; Reg_of_prg.1.
DR SMART: SM00315; RGS.1.
DR PROSITE: PS50132; RGS.1.
FT NON_TER 1
FT TER 85
SQ SEQUENCE 85 AA; 10108 MW; A96F32B2BB97D59 CRC64;

Query Match 0.8%; Score 7; DB 13; Length 85;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 NHVPAHD 46
Db 15 NHVPAHD 21

RESULT 33

```

```

09CLV3
ID 09CLV3 PRELIMINARY: PRT: 87 AA.
AC 09CLV3
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Hypothetical protein PM1099.
GN PM1099.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70.
RX MEDLINE=21145866; PubMed=11248100;
RA May B.-J., Zhang Q., Li L., Paustian M.L., Whitlam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL: AE006151; AAK03183.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 87 AA; 9829 MW; 5960BFA5BAAAC474 CRC64;

Query Match 0.8%; Score 7; DB 16; Length 87;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKRVLF 7
DB 1 MSKRVLF 7

RESULT 34
082102
ID 082102 PRELIMINARY: PRT: 88 AA.
AC 082102;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Putative flagellar regulatory protein.
GN YPO0720.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Tildall R.W., Holden M.T.G.,
RA Prentice M.B., Sebalha M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Farraga A.M.,
RA Chillingworth T., Cronin N., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001)
DR EMBL: AA414144; CAC89571.1; -
KW Flagella; Hypothetical protein; Complete proteome.
SQ SEQUENCE 88 AA; 9503 MW; DCCB35E838168073 CRC64;

Query Match 0.8%; Score 7; DB 16; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 533 AVARAIL 539
DB 77 AVARAIL 83

RESULT 35
09JSVO

```

```

ID 09JSVO PRELIMINARY: PRT: 91 AA.
AC 09JSVO;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Hypothetical protein NMA2123.
GN NMA2123.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=656599;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Fellwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491.";
RL Nature 404:502-506(2000).
DR EMBL: AL162758; CAB85336.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 91 AA; 10240 MW; 6327497E655F2DC0 CRC64;

Query Match 0.8%; Score 7; DB 16; Length 91;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LTPEDIQ 67
DB 8 LTPEDIQ 14

RESULT 36
098552
ID 098552 PRELIMINARY: PRT: 95 AA.
AC 098552;
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE A502L protein.
GN A502L.
OS Paramyxium bursaria chlorella virus 1 (PCV-1).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.
OX NCBI_TaxID=10506;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96400190; PubMed=8806566;
RA Kutish G.F., Li Y., Lu Z., Furuta M., Rock D.L., Van Etten J.L.;
RT "Analysis of 76 kb of the chlorella virus PCV-1 330-kb genome: map
RT positions 182 to 258.";
RL Virology 223:303-317(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20013326; PubMed=10544099;
RA Kaiser A., Vollmert M., Tholl D., Graves M.V., Gurnon J.R., Xing W.,
RA Lisee A.D., Nickerson K.W., Van Etten J.L.;
RT "Chlorella virus PCV-1 encodes a functional homosperrmidine
RT synthase.";
RL Virology 263:254-262(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20478054; PubMed=11021991;
RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;
RT "Characterization of a beta-1,3-glucanase encoded by chlorella virus
RT PCV-1.";
RL Virology 276:27-36(2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;

```

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RN [5] SEQUENCE FROM N.A.  
 RP Van Etten J.L.;  
 RA Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RL [6] SEQUENCE FROM N.A.  
 RP Van Etten J.L.;  
 RA Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RL [7] SEQUENCE FROM N.A.  
 RP Graves M.V., Van Etten J.L.;  
 RA Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 RL [8] SEQUENCE FROM N.A.  
 RP Graves M.V., Van Etten J.L.;  
 RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RL [9] SEQUENCE FROM N.A.  
 RP Gurnon J.R., Graves M.V., Van Etten J.L.;  
 RA Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: U42580; AAC96869.1; -  
 SQ SEQUENCE 95 AA; 11069 MW; 0811C752B1618A80 CRC64;

Query Match 0.88; Score 7; DB 12; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 817 QSTRGYA 823  
 Db 47 QSTRGYA 53

RESULT 37  
 O9UP45 PRELIMINARY; PRT; 98 AA.  
 ID 09UP45;  
 AC 09UP45;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE Type XIII collagen (Fragment).  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1] SEQUENCE FROM N.A.  
 RP MEDLINE=99357014; PubMed=10429945;  
 RA Kvist A.P., Latvanlehto A., Sund M., Horelli-Kuitunen N., Rehn M.,  
 RA Palotie A., Belter D., Pihlajaniemi T.;  
 RT "Complete exon-intron organization and chromosomal location of the  
 RT gene for mouse type XIII collagen (coll13a1) and comparison with its  
 RT human homologue."  
 RL Matrix Biol. 18:261-274(1999).  
 DR EMBL: AF071009; AAD48082.1; -  
 KW Collagen.  
 FT NON\_TER  
 SQ SEQUENCE 98 AA; 10228 MW; 93020572713BEE29 CRC64;

Query Match 0.88; Score 7; DB 4; Length 98;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 238 ALVAAA 244  
 Db 27 ALVAAA 33

RESULT 38  
 O96YV4 PRELIMINARY; PRT; 105 AA.  
 ID 096YV4;  
 AC 096YV4;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Hypothetical protein str2073.  
 GN str2073.  
 OS Sulfolobus tokodaii.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=111955;  
 RN [1] SEQUENCE FROM N.A.  
 RP STRAIN-JCM 10545 / 7;  
 RC PubMed=11572479;  
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
 RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,  
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
 RA Yoshizawa T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A.,  
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
 RA Oshima T., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic thermophilic  
 RT Crenarchaeon, Sulfolobus tokodaii strain7.";  
 RL DNA Res. 8:123-140(2001).  
 DR EMBL: AP000988; BAB67172.1; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 105 AA; 12168 MW; C2FC9B04E1AC83C CRC64;

Query Match 0.88; Score 7; DB 17; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 249 DIDSLII 255  
 Db 74 DIDSLII 80

RESULT 39  
 O92GXI PRELIMINARY; PRT; 119 AA.  
 ID 092GXI;  
 AC 092GXI;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE 50S ribosomal protein L22.  
 GN RPLV OR RCL1001.  
 OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsia.  
 OX NCBI\_TaxID=781;  
 RN [1] SEQUENCE FROM N.A.  
 RP STRAIN-MALISH 7;  
 RC MEDLINE=21442074; PubMed=11557893;  
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
 RA Raoult D.;  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";  
 RL Science 293:2093-2098(2001).  
 DR EMBL: AEO08652; AAL03539.1; -  
 DR InterPro: IPR001063; Ribosomal\_L22.  
 DR Pfam: PF00237; Ribosomal\_L22; 1.  
 DR ProDom: PD001032; Ribosomal\_L22; 1.  
 DR TIGRfams: TIGR01044; rplV\_bact; 1.  
 DR PROSITE: PS00464; RIBOSOMAL\_L22; UNKNOWN\_1.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 119 AA; 13253 MW; 5BE636CA907C7DA8 CRC64;

Query Match 0.88; Score 7; DB 16; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 491 KLNLYAA 497  
 Db 22 KLNLYAA 28

## RESULT 40

```

028742          PRELIMINARY;      PRT;      119 AA.
AC 028742;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein AF1530.
GN AF1530.
OS Archaeoglobus fulgidus.
OC Archaea: Euryarchaeota: Archaeoglobi: Archaeoglobales:
OC Archaeoglobaceae: Archaeoglobus.
OX NCBI_TaxId=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Relch C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Corton M.D., Spriggs T., Artiach P., Kalne B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RL reducing archaeon Archaeoglobus fulgidus.";
DR EMBL: AE000997; AAB89718.1; -.
DR TIGR: AF1530; -.
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 119 AA; 14025 MW; 7F75DA93E80C15C CRC64;

```

## Query March

Best Local Similarity 0.8%; Score 7; DB 17; Length 119;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LTPEDIQ 67

DB 102 LTPEDIQ 108

Search completed: April 28, 2003, 16:35:35  
Job time : 159 secs









CC production of BASB067 polypeptides and polynucleotides, for use  
 CC especially in therapeutic and prophylactic vaccines. It also  
 CC relates to methods for using such polypeptides and polynucleotides  
 CC in the prevention and treatment of microbial diseases, in diagnostic  
 CC assays for detecting diseases associated with microbial infections,  
 CC and assays for detecting expression or activity of BASB067  
 CC polypeptides or polynucleotides. Antibodies raised against  
 CC BASB067 can be used to treat humans with H. influenzae disease.

XX Sequence 578 AA:

Query Match 9.2%; Score 436.5; DB 21; Length 578;  
 Best Local Similarity 20.5%; Pred. No. 1.2e-21;

Matches 142; Conservative 105; Mismatches 235; Indels 211; Gaps 14;

OY 243 RAVGYDDIDLSIIRNS-IGEVYIIHDL--GEPYIDYRAVEREGADAKFTTVADEV 299  
 DB 73 RVFGYESSVFRERKOROGKRDLLIAHTVPEPTKIAGTDVQIEGAADENFNALRKNL 132  
 OY 300 PLLIGDVPHHGKYEETKKNLIENASAEHGFCDGRWLDKRSVDVILPONTADVSLITYDTGTQY 359  
 DB 133 P-KDGVLEHQTDDYKTAISRLALNRGYFDGNFKISRLSEIPETHQAWMMWMLFDSGVRY 191  
 OY 360 REDEVVFTIDPKTQNLTPDKLPVKRELEQLLVNMGAYNLQAVRALNDLIATRY 419  
 DB 192 HYGNTIF-----SHSQIRDDYLNINILNIKSGDPTLANNLSDLTSPFSSNW 237  
 OY 420 FNMVTEIVFPEREIOQNDQVSFEQSSSRPEPAQVDESTLEPIETVELTGILMDISP 479  
 DB 238 FSSV----- 241  
 OY 480 IEFASNLIOOKLNLVAAKARHLVMPDORVLAINHDDGVNRSIIIGRISDAVSAVARAIL 539  
 DB 242 -----LVOPRVN----- 248  
 OY 540 PDESENEVIDLPERPTALANRKTADYOSKVPYLVFVASDKPRGOGIGLGCWSPGTFL 599  
 DB 249 -----HKSQYDVEIILYPRKKMMELGCVSTGCGVHG 282  
 OY 600 VTKPEHNLINNDGYOAGAELELSEKKGVK-LYATKPLSHPLNDOLRATLGYOQEVFGHS 658  
 DB 283 QIGWKKPINSRGHSLRNLNLYLSAPKQTLKATYRMLPLKNPLNLYYDRAVGE---GEK 338  
 OY 659 TNGFDLSTRLEHEISRSIIIONGWNRTYSRLRDLKLTQAPPTWQDLPVDFVNGKPS 718  
 DB 339 EN-DTNRVLTLSLARKVWNAHGMQYFGCLRMKIDSF-TQAD----- 378  
 OY 719 OEALLAGVAHKTVAADNLVNPGRYR-----QVSLLEVSSGLVSDANMA 763  
 DB 379 -----ITDKTL---LLPYVGFTRTLRGSPATWGDQKITFDLSKRIWLSSESSI 427  
 OY 764 IARAGISGVYFSGONAVGNSRAHQMTGTIOAGYIWSDNFNVPRLRFPAGGDSIRGYA 823  
 DB 428 KVOASSAMVRYAEN-----HRYVARAEIGYLTGKIEKIPPLIRFPAGDSRVSRYG 480  
 OY 824 HDLSLPSIDKGYLTGGQVLAVGTAEVNEEFMKDLRLAVEGIGNAYDKGFTNDKIGAGV 883  
 DB 481 YKKIAPKRNCKLVGGSRLLTTSLEYOTQYVPPNMAATFADSGLAADYATKELRYGTGV 540  
 OY 884 GVRMASPVGOVRDVATGVKEGNPIKLHPEIG 916  
 DB 541 GVRMASPVGAIKFDIATPIRDKONSKNIQFYIG 573

RESULT 4

AA95821  
 ID AAY95821 standard; Protein; 576 AA.

AC AAY95821;

XX 07-NOV-2000 (first entry)

DE Haemophilus influenza nhl strain 289 BASB067 protein.

XX BASB067; outer membrane protein; antigen; vaccine; antibiotic;  
 KW antibacterial; screening; infection; diagnosis; therapy.  
 OS Haemophilus influenzae.

XX Key Location/Qualifiers  
 FH Peptide 1..20  
 FT /label= Signal--peptide  
 FT Protein 21..576  
 FT /label= Mature-protein

XX W0200047737-A1.

XX 17-AUG-2000.

XX 04-FEB-2000; 2000WO-EP00887.

XX 09-FEB-1999; 99GB-0002880.

XX (SMRK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX Ruelle J, Thonnard J;

XX WPT. 2000-515059/46.

XX N-PSDB: AAS0270.

PS BASB067 polypeptide and polynucleotide from Haemophilus influenzae are  
 used for diagnosing and treating H. influenzae infections -

Claim 1; Page 81-82; 87pp; English.

CC The present sequence is that of outer membrane protein BASB067 of  
 CC Haemophilus influenzae non-typable (Nth) strain 289. BASB067 is  
 CC a surface expressed protein that is recognised by the immune system.  
 CC It shows homology to the protective surface antigen D15 of  
 CC H. influenzae and has a similar secondary structure. Its  
 CC N-terminal domain is predicted to contain a mixture of alpha-helix  
 CC and beta-strands, and could be used as a vaccine antigen. The  
 CC C-terminal domain is predicted to form a beta-barrel composed of  
 CC anti-parallel, amphipathic beta-strands. The external loops of  
 CC the beta-barrels of integral outer membrane proteins frequently  
 CC contain immunodominant B-cell epitopes, making the C-terminal  
 CC domain of BASB067 a strong candidate vaccine antigen. The  
 CC invention relates to recombinant materials and methods for the  
 CC production of BASB067 polypeptides and polynucleotides, for use  
 CC especially in therapeutic and prophylactic vaccines. It also  
 CC relates to methods for using such polypeptides and polynucleotides  
 CC in the prevention and treatment of microbial diseases, in diagnostic  
 CC assays for detecting diseases associated with microbial infections,  
 CC and assays for detecting expression or activity of BASB067  
 CC polypeptides or polynucleotides. Antibodies raised against  
 CC the mature portion of this BASB067 polypeptide can be used to treat  
 CC humans with H. influenzae disease.

XX Sequence 576 AA:

Query Match 9.0%; Score 427.5; DB 21; Length 576;  
 Best Local Similarity 20.3%; Pred. No. 5e-21;

Matches 141; Conservative 106; Mismatches 235; Indels 211; Gaps 14;

OY 243 RAVGYDDIDLSIIRNS-IGEVYIIHDL--GEPYIDYRAVEREGADAKFTTVADEV 299  
 DB 71 RVFGYESSVFRERKOROGKRDLLIAHTVPEPTKIAGTDVQIEGAADENFNALRKNL 130  
 OY 300 PLLIGDVPHHGKYEETKKNLIENASAEHGFCDGRWLDKRSVDVILPONTADVSLITYDTGTQY 359  
 DB 131 P-KDGVLEHQTDDYKTAISRLALNRGYFDGNFKISRLSEIPETHQAWMMWMLFDSGVRY 189  
 OY 360 REDEVVFTIDPKTQNLTPDKLPVKRELEQLLVNMGAYNLQAVRALNDLIATRY 419  
 DB 190 HYGNTIF-----SHSQIRDDYLNINILNIKSGDPTLANNLSDLTSPFSSNW 235

```
QY 420 FNNVNTLIVPEREIOINDVQSFEOSSSKTERPAQVDESTLEPIVETVELTDLGIMDISP 479
|:|
Db 236 FSSV----- 239
QY 480 IEFASNLIDOKLNLVAKARHLVMPDDRYLAINHDDGVNRSILGRISDAVAVARAIL 539
|:|
Db 240 -----LVQPNV----- 246
QY 540 PDESENEVIDLPERTALANRKPADVYQSKKVPYVVASDKPRDGOIGLGWSDTGTRL 599
|:|
Db 247 -----HKSKTVDEIILLPYRKKNAMELGCFATDGVHG 280
QY 600 VTKFEHNLINRDYQQAQAEIRLSDKKGVK-LYATKPLSHPLNDQLRATGLGYQGEVGH 658
|:|
Db 281 QIGETKPMINSRGLSRSLNLYLAPKQTLKATYMPPLKNPLNLYYFVAVGWE-----GEK 336
QY 659 TNGFDLSTRTLEHEISRSIIONGGMNRTYSLRYRLDKLTKQAPETQODLPVDVNGKPS 718
|:|
Db 337 EN--DTMTRYLTLTSLALRYNNNAHGMQYFGGLRTKIDSE--TQAD----- 376
QY 719 OEALLAGVAHVKTVADNLVNPMDGR-----QRTSLGVSSGLVSDANMA 763
|:|
Db 377 -----ITDKTL-----LXPTVGFTRTLRCGSEFATMGDQKITPDLKRIMLSESSFI 425
QY 764 IARAGISGVYSEFGDNVAGSNRAHOMTGIQAGIYMSDNFNHVPYRLRPFAGDOSINGYA 823
|:|
Db 426 KYQASSAMVRYTAEN-----HRIVARAEIGYLHTKIDIEKIPPLTRFFAGGDRSVG 478
QY 824 HDLSLPSIDGYLTGGVAVLAVGTAEVYEFMKDLRLAVFGIGNAVYKGFNDOKIGAGV 883
|:|
Db 479 YKRIAPKRNNGKLVGSRLLTTSLENOYVYPMWMAATFADSGLAUDNYTAKELRYGTGV 538
QY 884 GVRMASPVGVRVDVATVKEEGNPILKHEFFIG 916
|:|
Db 539 GVRMASPVGAIKFEDIATPIRDKMSKNIGFYIG 571
RESULT 5
AAU03958
ID AAU03958 standard; Protein; 792 AA.
XX
AC AAU03958;
XX
DT 23-OCT-2001 (first entry)
XX
DE Neisseria meningitidis serogroup A antigenic protein #1.
XX
KM Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;
XX
KM bacterial infection; baculovirus; yeast.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT Protein /note="Signal peptide"
FT Protein 22..792
FT Protein /note="Mature N. meningitidis serogroup A antigen"
XX
PN WO200138350-A2.
XX
PD 31-MAY-2001.
XX
PF 28-NOV-2000; 2000WO-IB01851.
XX
PR 29-NOV-1999; 99GB-0028197.
XX
PR 09-MAR-2000; 2000GB-0005698.
XX
PA (CHIR-) CHIRON SPA.
PA (STAT-) STATENS INST FOLKEHELS.
XX
PI Giuliani MM, Pizsa M, Rappuoli R, Holst J.
XX
DR WPI: 2001-381289/40.
```

```
DR N-PSDB: AAS07278.
XX
PT Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
PT gonorrhoeae, useful in the manufacture of a medicament for treating and
PT preventing Neisseria bacteria infection.
XX
PS Claim 1; Page 66-68; 92pp; English.
XX
CC The sequence represents a Neisseria meningitidis serogroup A 85 kDa
CC antigenic protein. Neisseria meningitidis colonises the pharynx, causing
CC meningitis and, occasionally, septicaemia in the absence of meningitis.
CC This antigenic protein is useful in the manufacture of a medicament for
CC treating or preventing infection due to Neisseria bacteria, such as
CC meningitis and septicaemia. It is also useful as a diagnostic reagent for
CC detecting the presence of Neisseria bacteria or antibodies raised against
CC Neisseria, and as a reagent for raising the antibodies. The Neisseria
CC nucleotide sequences can be expressed in a variety of different
CC expression systems, for example, mammalian cells, baculoviruses, plants,
CC bacteria and yeast.
CC Note: There are two versions of this sequence displayed in the
CC specification (see AAU04451).
XX
SQ Sequence 792 AA;
Query Match 5.6%; Score 265.5; DB 22; Length 792;
Best Local Similarity 20.9%; Pred. No. 1.1e-09;
Matches 192; Conservative 119; Mismatches 352; Indels 255; Gaps 39;
QY 116 TPLSLEELFAOESTREMGINPND-----YIPYQGEQPNSEVVPVPLEPKRLKRLY 169
|:|
Db 16 SPLAFADFTIODIRVEGLORTPESTVFNYLIPVKGDYRN-----DTGSAIKSLY 66
QY 170 ARLENDGVNKP---RLKAKFYQSSQSGENSAISSHOKTEPYANIKALED--ITQESA 224
|:|
Db 67 ATGFFDDVRYETADGQLLLVYIERPTIGSLNITAKMLQND-----IKKNLESFGLAOSQY 123
QY 225 MD---LNGSIPRLRQ-----TALVAVARAVGYIDLSIIRNSIGEVDVYIIDL 269
|:|
Db 124 FNGATLNOAVAGLKEEYLGSGKLNLIQTTPKYTKLARNRVDRIDITDECKSAKITDIEFE- 182
QY 270 GEPIYIDYRAVEVRGEGADDAKFTTVADEVPLIGDVFHHGKYETKKNLJENASAEHGYF 329
|:|
Db 183 GNOVYSDRKILMR-----QMSLTEGGIMTWLTRSDFRQKFAQDMEKYVDFYQNNGYF 235
QY 330 DGRMLDSVDVILPNDADVSLIYDTGTYRFDEYVFFETIDPKNOLTTDPDKIPVREL 389
|:|
Db 236 DFRLLDTIDIONEDKTRQTKITIVHGGREFRWKV---SLEGDTNE-----VPKAE 283
QY 390 LEOLTYVNMGEAVNLQAVRALSNDLIATRYFNMVNTEIVPEREIOINDVQSFEOSSSR 449
|:|
Db 284 LEKLITPKKPKWYERQCMTVLQ-----EIQNRKGSAGYATIS-- 320
QY 450 TEPAQVDESTLEPIVETVELTDLGIMDISP-----IEFSASNLIDOKLNLVAKARH 501
|:|
Db 321 -----EISVQPLPNAGTQTVDFVLIIEPRKIYVNEIHTGNNKTRDE--VYRRELRLQ 371
QY 502 LYDMPD-----RYLAINHDDGVNRSILGRISDAVAVARAILPDESENEVIDLP 552
|:|
Db 372 MESAPYDTSKLORSKEVELLGFEDNVQ-----FDVAVPLAGTPDK-----VDL- 414
QY 553 RTALANRKPADVYQSKKVPYVVASDKPRDGOIGL--GWSGDTGRLTKFEHNLINR 610
|:|
Db 415 NMSLTERST-----GSLDISAGWVDGT--LYMSAGVSODNL 449
QY 611 DGQAQAEIRLSDKKGVKLYATKPLSHPLNDQLRATGLGYQGEVGHSTNGFDLSTRTLE 670
|:|
Db 450 FGTGKSAALRASRSK--TTLNGSLSTFDPYFTADGVSLGY--DIYGAFDPRAASTSVKQ 505
QY 671 HEISRSIIONGGMNRTYSLRYRLDKLTKQAPPE-----TWQDLP---VDFVN--GK-PSQ 719
|:|
Db 506 YKTTTA-----GGGVRMGIPVTEYDRVNFGLAAEHLTVNVTAKPRKRYADFIIRKYGTGDA 561
QY 720 EALLAGVAHVKTV-----ADNLVNPMDGRYRTSLGVSSGLVSDANMAIARAGISGVY 773
```

Db 562 DGSFGLLYKGVGWRKTKDASWPTRGY-----LGVNAEIALPESKIQY 608  
QY 774 SEGDNAVGSNRAHQMTG-----GIOAGYIMSDNFHVPYRLRFPAGDQ 817  
Db 609 -----YSATHNQTWTFPLSKFTTLMLGGEVGIAGY---GRKKEIPFEFNFGGGLG 657  
QY 818 SIRGAHDSLSPIISKGYLTGGQVLAAG---TAENYEFM-----KDLRLAVFGDI 865  
Db 658 SVRGYESGTLGP---KYDEVGEKISYGNKKANYSABELLFPMPGAKDARTVRLSLFADA 714  
QY 866 G-----NAYDKGTNDTKIGAGVGYRMASPVGOVVDVATGVK 903  
Db 715 GSYWDGRTYTAENGNKNSVYSENHAKSTFTNELKYSAGAVTWLSPUGPMKFTAYPLK 774  
QY 904 E--EGNPIKLHFFICTPF 919  
Db 775 KKPEDEIQRFQFQLGTTF 792

## RESULT 6

AAB84745  
ID AAB84745 standard; Protein; 792 AA.

AC AAB84745;

DT 17-SEP-2001 (first entry)

DE Amino acid sequence of a Neisseria gonorrhoeae protein.

KW Serogroup B protein; outer membrane protein; Neisserial infection;  
vaccine.

OS Neisseria gonorrhoeae.

XX Key Location/Qualifiers

FT Peptide 1..21 /note= "signal peptide"

FT Protein 22..792 /note= "mature protein"

XX W0200152885-A1.

XX 26-JUL-2001.

PD 17-JAN-2001; 2001MO-IB00166.

PR 17-JAN-2000; 2000GB-0001067.

XX 09-MAR-2000; 2000GB-0005699.

PA (CHIR-) CHIRON SPA.

XX Pizsa M, Rappuoli R, Giuliani M;

DR WPI; 2001-451895/48.

XX N-PSDB; AAH42129.

PT Composition for treating or preventing infection to, detecting, or for  
PT raising antibodies against Neisserial bacteria, comprises an N.  
PT meningitidis serogroup B outer membrane preparation and an immunogenic  
PT component -  
XX  
XX Disclosure; Page 65-67; 83pp; English.

XX The present sequence represents a Neisseria gonorrhoeae protein. The  
CC protein is used to produce the compositions of the invention. The  
CC specification describes a composition, comprising a Neisseria  
CC meningitidis serogroup B outer membrane preparation and an immunogenic  
CC component. The immunogenic component is protein disclosed in W099/57280,  
CC W099/36544, W099/24578, W099/66791, W097/28273, W096/229412, W095/03413,  
CC W099/31132, W099/58683, W099/55873, and/or N. meningitidis protein P07A,  
CC T07A, T07B, P11C, OPA, or Omp85. The composition is used for making  
a medicament for treating or preventing infection due to Neisserial

CC bacteria; a diagnostic reagent for detecting the presence of Neisserial  
CC bacteria or of antibodies raised against Neisserial bacteria; and/or  
CC a reagent which can raise antibodies against Neisserial bacteria. It may  
CC also be used as a vaccine.

SQ Sequence 792 AA:

Query Match 5.6%; Score 265.5; DB 22; Length 792;

Best Local Similarity 20.9%; Pred. No. 1.1e-09;

Matches 192; Conservative 119; Mismatches 352; Indels 255; Gaps 39;

QY 116 TPLSLEFLAQSSTENGIPND-----YIPXYOGQPMSEVVVPPTLPKGLIKLY 169  
Db 16 SPLAADFPTIIDIIRVEGLQRTPESTVFNYLPVKYGDYV-----DTHGSAIIKSLY 66  
QY 170 ARLFNDGVNKPV---RLKAKFYSSOSGETSAIGSSHOKEPEYANIKALED---ITQESA 224  
Db 67 ATGFDDVAVETADGOLLTVIERPTIGSLNITGAKMLONDA---TKNLESFGLAOSQY 123  
QY 225 MD---LNGSIPRLRQ-----TALVARAVGYIDLSIRNSIGEVYIIDL 269  
Db 124 FNOATLNOAVAGLKEEYILGRKLNIOITPKYTKLARNRVIDITIDEKSAKITDIEFE- 182  
QY 270 GEPYVIDYRAVEVRBEGADKAFITVADDEVPLLIGVFHNGKYEKKNLLENASAHEGYF 329  
Db 183 GNOVYSDRKLNR-----QMSLTBEGITWTWLTFRSDREBQKFAODMEKVTDFYONNGYF 235  
QY 330 DGRWLDRSYDVILPNTADVSLIYDTGQYRFDEVVFTIDRTNQLTDPDKLPKREL 389  
Db 236 DFRILDTIQTNEDKTRQIKITVEHGGFRMGKV---SIBEDTNE-----VPRKAE 283  
QY 390 LEQLITVMKGEYNIQAVRALSNDLIATRFMMVNTETVFPEREOIQNDQVFEQSSSR 449  
Db 284 LEKLLTMKRGKWEQOMTAVLG-----ETONMGSGAGYAYS-- 320  
QY 450 TPAQVDSLEPTEVETVELTGGIIMDISP-----IEFSASNLIODKMLUVAKARH 501  
Db 321 -----ELSVQPLRNAGIKTYDFVLIHEPGRKIYVNEHITGNKKTDE--VIRELRO 371  
QY 502 LYDMPDD-----RYLAINHDDGVNRSITGRISDAVAVARAILPDESENEYIDLP 552  
Db 372 MESAPYDTSKLDQSKEREVELLGYFDNVQ-----FDAVPLAGTDPK----VDL-- 414  
QY 553 RTALANRKTTPADYVOSKVPVLYVFAASDKPRDQIGL--GMSDGTGTRLYTFEHLNLR 610  
Db 415 NMSLTERST-----GSLDLSAGWQDTG--LWMSAGVSQDNL 449  
QY 611 DGYQAGAEIRLSEDDKGVKLYATKPLSHPLNDQLRATIGOOEFGHSTNGFDLSTRILE 670  
Db 450 FGTGKSALRLASRSK--TTLNGSLSFPTDPTADGVSLGY--DIYGKAFDPKAKASTYKQ 505  
QY 671 HEISRSITONGGMNFTYSRLRYRLDKLKTQAPPE---TWQDLP--VDENV--GK-PSQ 719  
Db 506 YKTTTA---GGGVAMGIPVTEYDRVNFGLAEHLTVMTYNNAPRYADFIKKYTGCA 561  
QY 720 EALLAGVAVHKTIV-----ADNLVPMRGYRQRYRLEEVSSGLVSDAMATARAISGVY 773  
Db 562 DGSFGLLYKGVGWRKTKDASWPTRGY-----LGVNAEIALPESKIQY 608  
QY 774 SEGDNAVGSNRAHQMTG-----GIOAGYIMSDNFHVPYRLRFPAGDQ 817  
Db 609 -----YSATHNQTWTFPLSKFTTLMLGGEVGIAGY---GRKKEIPFEFNFGGGLG 657  
QY 818 SIRGAHDSLSPIISKGYLTGGQVLAAG---TAENYEFM-----KDLRLAVFGDI 865  
Db 658 SVRGYESGTLGP---KYDEVGEKISYGNKKANYSABELLFPMPGAKDARTVRLSLFADA 714  
QY 866 G-----NAYDKGTNDTKIGAGVGYRMASPVGOVVDVATGVK 903  
Db 715 GSYWDGRTYTAENGNKNSVYSENHAKSTFTNELKYSAGAVTWLSPUGPMKFTAYPLK 774  
QY 904 E--EGNPIKLHFFICTPF 919









Best Local Similarity 20.5%; Pred. No. 3.2e-09;  
Matches 191; Conservative 121; Mismatches 342; Indels 278; Gaps 40;

```

OY 116 TPULSELEFAQESTENGINPND-----YIPEYQGPNSSEVVVPTLPPEKGLIKRLY 169
DB 16 SPLALADFTIODIRVEGLQRTPESTVFNYLPVKGDTYN-----DTHGSAITIKSLY 66
OY 170 ARLEFNGCVNKKVP---RLNKKFYOSSGSETSAIGSSHOKTEPYANIKALED--ITQESA 224
DB 67 ATGFEDDVAVETADQGLLTVIERPTIGSLNTGAKMLQND--IKKNLESFGLAQSOY 123
OY 225 MD---LNGSIPRLQ-----TALVARAVGYDIDLSITRNSIGEVDVITIDL 269
DB 124 FNOATLNQAVAGLKEEYLGKGLNTIQTIPKVTKLARNRNDIDITIDEKSAKITDIEFE- 182
OY 270 GEPVYIDYRAVEVRGEGADDKAFTVADEVPLLIGDV-----FHNGKETKKNLIE 320
DB 183 GNOVYSDRKLMR-----QMSLTGEGITWTLTRSNQFNQCKFAQDMEKVT 226
OY 321 NNSAEHGYDFGRMLDRSDVILPDMTADVSLIYDTQYRFDEVFETIDPKTNOLTTDP 380
DB 227 DTYQNNGYDFPILDDIOTNEDKTKOTIKITVHESGFRFGKV---SIEGDINE----- 278
OY 381 DKLPVKRELLEQLTYNMGEAYNLQAVRALSNDLATRYFNMYNTEIVPEREQIONQV 440
DB 279 ---VPKAELEKLTWPKPKWYERQOMTAVLG-----EIONRMG 313
OY 441 SFEQSSSTTEPAQVDESLFEVIEVELTDGLMDISP-----IEFSASNLIQDKL 492
DB 314 SAGVAYVS-----EISVQPLPNAETKVDFVLHIEPGRKIYVNEIHITGNKKRDE- 363
OY 493 NLVAAKARHLVMPDD-----RVLAINHDDGVNRSILGRISDAVAVARAILPDES 543
DB 364 -VVRRELQMEAPFDTSKLORSKEVELLGYFDNVQ-----FDAVPLAGTPDK- 411
OY 544 ENEVIDLPERTALANRKTTPADVYOSKKVPLVYFVASDKPRDQIGL--GMSDPTGLRYT 601
DB 412 ---VDL--NMSLTERST-----GSLDLSAQMVDTG--LVM 440
OY 602 KFEHLINRDQQAQAEELRLSDKGVKIYATKPLSHPLNDQLRATLGYQOEVEFGHSTNG 661
DB 441 SAGVSQDMLFTGKSAALRASRSK--TTLNGSLSFTDPFTFADGVSLGY--DVGKAFDP 496
OY 662 FDLSTRLEHETSRIIONGWNR-----TYSLRYLRLDKLQOAPPEWQDLP---VDFV 713
DB 497 RKASISIQYKTTTA---GAGIRMSVPEYEDRVNFGCLVAEHLTVNTYNKAPKHVADFI 552
OY 714 N-GK-PSQEALLAGVAVHKTV-----ADNLVPMRGYRORYSLEVSSGLVSDANMAI 764
DB 553 KRYGKTDTGDSFGKWLKYGTVGMGRNKTDALMPTRGY-----LTCVNAEI 599
OY 765 ARAGISGYVSGDNAYGSRNHHQMTG-----GIOAGYIWSNENHNPYR 808
DB 600 ALPGSKLOY-----YSATHNQTFEPLSKTFTLMLGGEVINGY--GRTKEIPFF 648
OY 809 LRFAGGDSIRGTAHSLSPISDKGYLTGGCVLAVG---TAEVYVEEM-----KD 856
DB 649 ENFYGGGIGSRYGESGLGP---KYVDEYGEKISYGGNKKANSAELLFPFGKAKDAFT 705
OY 857 LRLAVFDGICNAV-----KGFNTDKIGAGVGVRRAS 889
DB 706 VRLSLFADAGSVMDGKTYDDNSSATGGRVONIGAGNTHKSTFNNELRYSAGAGVATMLS 765
OY 890 PGQVNRDVATGVKE--EGNPRIKLHFIQTPE 919
DB 766 PLGPMKFSYAVPLKKKPEDEIQRFQFOLGTF 797

```

RESULT 11  
AAU04451  
ID AAU04451 standard; Protein; 797 AA.  
XX  
AC AAU04451;

```

XX 23-OCF-2001 (first entry)
XX
XX Neisseria meningitidis serogroup A antigenic protein #2.
DE
XX
XX Serogroup B antigen; pharynx; meningitis; septicemia; mammalian cell;
KW bacterial infection; baculovirus; yeast.
XX
XX Neisseria meningitidis.
XX
XX
XX Key location/Qualifiers
FH Peptide 1..21
FT /note="Signal peptide"
FT Protein 22..797
FT /note="Mature N. meningitidis serogroup A antigen"
XX
XX W0200138350-A2.
XX
XX 31-MAY-2001.
XX
XX 28-NOV-2000; 2000WO-IB01851.
XX
XX 29-NOV-1999; 99GB-0028197.
XX 09-MAR-2000; 2000GB-0005698.
XX
XX (CHIR-) CHIRON SPA.
XX (STAR-) STARENS INST. FOLKEHELESE.
XX
XX Giuliani MM, Pizsa M, Rappuoli R, Holst J;
XX WPI; 2001-381289/40.
XX
XX Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
PT gonorrhoeae, useful in the manufacture of a medicament for treating and
PT preventing Neisserial bacterial infection.
XX
XX Claim 1; Page 39-40; 92pp; English.
XX
XX
XX The sequence represents a Neisseria meningitidis serogroup A 85 kDa
CC antigenic protein. Neisseria meningitidis colonises the pharynx, causing
CC meningitis and, occasionally, septicemia in the absence of meningitis.
CC This antigenic protein is useful in the manufacture of a medicament for
CC treating or preventing infection due to Neisseria bacteria, such as
CC meningitis and septicemia. It is also useful as a diagnostic reagent for
CC detecting the presence of Neisseria bacteria or antibodies raised against
CC Neisseria, and as a reagent for raising the antibodies. The Neisserial
CC nucleotide sequences can be expressed in a variety of different
CC expression systems, for example, mammalian cells, baculoviruses, plants,
CC bacteria and yeast.
CC Note: there are two versions of this sequence displayed in the
CC specification (see AAU03958).
XX
XX
XX Sequence 797 AA:

```

Query Match 5.5%; Score 259; DB 22; Length 797;  
Best Local Similarity 20.5%; Pred. No. 3.2e-09;  
Matches 191; Conservative 121; Mismatches 342; Indels 278; Gaps 40;

```

OY 116 TPULSELEFAQESTENGINPND-----YIPEYQGPNSSEVVVPTLPPEKGLIKRLY 169
DB 16 SPLALADFTIODIRVEGLQRTPESTVFNYLPVKGDTYN-----DTHGSAITIKSLY 66
OY 170 ARLEFNGCVNKKVP---RLNKKFYOSSGSETSAIGSSHOKTEPYANIKALED--ITQESA 224
DB 67 ATGFEDDVAVETADQGLLTVIERPTIGSLNTGAKMLQND--IKKNLESFGLAQSOY 123
OY 225 MD---LNGSIPRLQ-----TALVARAVGYDIDLSITRNSIGEVDVITIDL 269
DB 124 FNOATLNQAVAGLKEEYLGKGLNTIQTIPKVTKLARNRNDIDITIDEKSAKITDIEFE- 182
OY 270 GEPVYIDYRAVEVRGEGADDKAFTVADEVPLLIGDV-----FHNGKETKKNLIE 320
DB 183 GNOVYSDRKLMR-----QMSLTGEGITWTLTRSNQFNQCKFAQDMEKVT 226

```

QY	321	NASAHGYFDGDMKIDRSVDVILPNDTADVLSLYTGTQYRFRDEVFTIDBKTNOLTDP	360
Db	227	DFYQNGGFEDRILDTDITQTNEDKTKQTIKTIHVEGGFRFGKVA--SISGDNTNE-----	278
QY	381	DKLPRKRELLDQLTVNNGEAYNLQAVALSNDLILAGRYFMVMTETLVPFEREQINDQY	440
Db	279	-----VPKAELEKLLTKRKGGKTKERQOMTAVLG-----ETQNRMG	313
QY	441	SFEQSSSRTERPAQVDESLTEPVLETVELTDGILMDISP-----IEFSASNLIODKL	492
Db	314	SAGYAYS-----EISVQPLPNAETKTYVDPLHIEPRKTIYVNEIHTGNKKTKPDE-	363
QY	433	NIUVAKAKHLKIDMPD-----RVLIATHDQGVNRSLIGRTSDAVSAVARAILPDES	543
Db	364	-VVRRELLQOMESAPRTDTSKIDRSKERVELLGYPDNVQ-----FDVAPLAGTPDK-	411
QY	544	ENEVDLDEPFRALANRKTPAVDYQSKVPLVFAVASDKPRDQIGL--GGAQSDPTPTLYT	601
Db	412	----VDL--NMSLIERST-----GSIDLSAGWVQDTG--LVM	440
QY	602	KFEHNLINRDGYQAGAEILRLSEDKKGVLYATKPELSPHLNDOLRATLGGYQOEFGHSTNG	661
Db	441	SAGYQODMLFTGKSAALRASRSK--TTLNGSLSTPTDPTPFADGVSLSGY--DVYGAFRD	466
QY	662	FDLSRLTLEHEHSKSIIONGGMNR-----TYSLKYRLDKLKTQAPRPTWODLP---VDVY	713
Db	497	RKASITSIKQYTTTA---GAGIRMSVPVPEYEDRVNFGLVAEHLTVNTYKKAPKHVADFI	552
QY	714	N--GK--PQEQALLAGVAHVKTIV-----ADNLVMPMGYRQRYSLKESGSGSLVSDAMAI	764
Db	553	KRYGTTDSDTSFGKMLYKGTGVGGRKNTDSALMPTKGY-----LTGVAAEI	599
QY	765	ARAGISGVYSEGDNAYGSGNRAHQMTG-----GIAQGYIMSDNFNNHPYR	808
Db	600	ALPGSKLQY-----YSATHNQTWEPFLSKTFTLMLGGEVGIAGY--GRTKELRPF	648
QY	809	LRFKFGQSGISRGYAHBSLSPISDKGYLTGGQVLAVG---TAETNYEFM-----KD	856
Db	649	ENFYGGGGLGSYRGYESGTLGP--KYVDEYGEKISYGNKKANVASAELLPPMPGAKDART	705
QY	857	LRLAVFGDIGNAYD-----KGFTNDTKIGAGVGYRWAS	889
Db	706	VRLSLFDAGSVMOGKTYTDNDSSSATGGRVONITGACNTHKSTFTTNELRISAGAGVATWLS	765
QY	890	PVGQVRVDVATGKRE--EGNDIKLHFFIGRPF	919
Db	766	PLGPMKFSYAYPLKKKPEDEIQRFQGLGTF	797
RESULT 12			
AAB84744			
ID AAB84744 standard; Protein: 797 AA.			
AAB84744:			
XX	17-SEP-2001	(first entry)	
XX	Amino acid sequence of a Neisseria serogroup B protein.		
XX	Serogroup B protein; outer membrane protein; Neisserial infection;		
XX	Neisseria meningitidis.		
XX	Key	Location/Qualifiers	
XX	Peptide	1..21	
XX	Protein	/note="signal peptide"	
XX		22..797	
XX		/note="mature protein"	
XX	WO200152885-A1.		

PD 26-JUL-2001.

PF 17-JAN-2001: 2001WO-IB00166.

XX 17-JAN-2000: 2000GB-0001067.

XX 09-MAR-2000: 2000GB-0005699.

XX (CHIR-) CHIRON SPA.

XX Pizza M, Rappuoli R, Giuliani M;

XX WPI: 2001-451895/48.

DR N-PSDB: AAH42128.

PT Composition for treating or preventing infection to, detecting, or for

PT raising antibodies against Neisserial bacteria, comprises an N.

PT meningitidis serogroup B outer membrane preparation and an immunogenic

PT component -

PS Disclosure: Page 59-61; 83pp; English.

XX

XX The present sequence represents a Neisseria serogroup B protein. The

CC protein is used to produce the compositions of the invention. The

CC specification describes a composition, comprising a Neisseria

CC meningitidis serogroup B outer membrane preparation and an immunogenic

CC component. The immunogenic component is protein disclosed in WO99/57280,

CC WO99/36544, WO99/24578, WO99/65791, WO97/28273, WO96/29412, WO95/03413,

CC WO99/31132, WO99/58663, WO99/55873, and/or N. meningitidis protein PorA,

CC TbpA, TbpB, PilC, Opa, or Omp85. The composition is used for making

CC a medicament for treating or preventing infection due to Neisserial

CC bacteria; a diagnostic reagent for detecting the presence of Neisserial

CC bacteria or of antibodies raised against Neisserial bacteria; and/or

CC a reagent which can raise antibodies against Neisserial bacteria. It may

CC also be used as a vaccine.

CC

XX

Sequence 797 AA:

Query Match 5.5%: Score 259; DB 22; Length 797;

Best Local Similarity 20.5%: Pred No. 3.2e-09;

Matches 191; Conservative 121; Mismatches 342; Indels 278; Gaps 40

116 TPLSEELTFAOSTEMGIMPND-----YIPEYOGBOPSEVVVPPTLPEPKGLIKRLY 169

Db 16 SPLADLFTIQDIRVEGLQRTPESTVFENYLPVKVGDTYN-----DTHGSAIRIKSLY 66

QY 170 ALPLFDGVAKVP--RLAKKTYQSSSGETSAIGSSHQKTEPANKALED--ITQESA 224

Db 67 AAGFFDDVAVETADQQLLTVTIERPTIGSLNTGAKMLQND--TKNNLESGLAOSQY 123

QY 225 MD--LNGSIPRLRQ-----PLVAVARAAGYVDIDSLIRNSIGEDVVIHDL 269

Db 124 FNGATLNLQAVAKLKEEYLGKRLNIQITPKYVTKLARNRDIDITIDEGSAKTIIDLEFE 182

QY 270 GEPVYIDYRAVEVRGEGADDKAFTTVADEVPLLIGDY-----FHNGKYETKKNLIE 320

Db 183 GNOVYSDRKLMR-----QMSLTGEGIMTWLIRSNQFNQKRAQADEKRYT 226

QY 321 NASAEHGFGDGMILRSVDVILPDMTADVSLIYDGTQVRFPEVYEFITDPTNQLTTPR 380

Db 227 DTYQNNNGTFDFPILDTDIOTNEDKTKQTKITVHBEGRFRKRV--STEGDTNE----- 278

QY 381 DKLPAVKRELLQDLTLVNMGEAYNLQAVRALSNDLIATRFNMVNTIEIVPEREQIONDQV 440

Db 279 ----VPKALEKLEKLLIMKPGKWEYERQGMNAVILG-----EIQNRMG 313

QY 441 SFEQSSSSRTPEAQVDESLPEVIVETVELTGDILMDISP-----IEFSASNLIQDKL 492

Db 314 SAGAYVS-----EISVQPLPNAETKTYDFVLHIEGKRIYVNEIHITNNKTRDE- 363

QY 493 NLVAKARHLLDMPD-----RVLAIHNDQDVNSILGRISDAVASAVARALPDPS 543

Db 364 -VVRRELROMESAPYDTSKLQSRKRVELLGTFDNYO-----FDAYVLAAGTPDK- 411



```

Db      600 ALPGSKLQY-----YSATHNQTFPPFLSKFTTLMLGGEVGIAGY---GRTKEIPFF 648
OY      809 LRFEPAGDOSIRGYAHDSLPISDKGYLTGGQYLAVG---TAEYNYEFM-----KD 856
Db      649 ENFYGGGLGSVRYEGSGTIGP---KYDEYGEKISYGNKKANVASALLEPMPGAKDART 705
OY      857 LRLAVFGDIGNAYD-----KGFTNDTKIGAGVGVWAS 889
Db      706 VRLSLFLFADAGSVWDGKTYDDNSSSATGGRVQNTYGAGNTHKSTFTNRLRYSGAGAVTWLS 765
OY      890 PYGQVRVDVATGVKE---EGNPILKHFFIGTFP 919
Db      766 PLGPMKFSTAYPLKKRPEDEIQRFQGLGTTTF 797

RESULT 14
AAB23784
ID      AAB23784 standard; Protein; 797 AA.
XX
AC      AAB23784;
XX
DT      12-JAN-2001 (first entry)
XX
DE      Neisseria meningitidis serogroup B amino acid sequence.
XX
KW      CG motif; complete Freund's adjuvant; phosphorothioate; immunogenic;
KW      Neisseria antigen; Neisseria meningitidis; Neisseria gonorrhoeae;
KW      bactericidal; antibacterial; vaccine; immunostimulatory; infection;
KW      immune response.
XX
OS      Neisseria meningitidis.
XX
PN      WO20050075-A2.
XX
PD      31-AUG-2000.
XX
PF      09-FEB-2000; 2000WO-IB00176.
XX
PR      26-FEB-1999; 99US-0121792.
XX
PA      (CHIR-) CHIRON SPA.
XX
PI      Grandi G, Rappuoli R, Giuliani MM, Pizza M;
XX
DR      WPI; 2001-015529/02.
XX
PT      Immunogenic composition useful for stimulating an immune response in a
PT      mammal against Neisseria infection, comprises Neisseria antigen and an
PT      adjuvant composition comprising an oligonucleotide with a CG motif -
XX
PS      Claim 22; Page 32; 39pp; English.
XX
CC      The present invention describes an immunogenic composition (I)
CC      comprising a Neisseria antigen and an adjuvant composition comprising an
CC      oligonucleotide comprising at least 1 CG motif. Also described is an
CC      adjuvant composition (II) comprising an oligonucleotide which comprises
CC      at least 1 CG motif and a complete Freund's adjuvant (CFA), where the
CC      oligonucleotide preferably comprises at least one phosphorothioate bond.
CC      AA92339 to AA92385 represent specifically claimed oligonucleotides of
CC      the present invention. (I) is useful for stimulating an immune response
CC      in a mammal, preferably a human, against Neisseria infection, preferably
CC      Neisseria meningitidis infection, and in the manufacture of a medicament
CC      for inducing a protective immune response in a mammal. The present
CC      sequence represents the claimed Neisseria meningitidis serogroup B amino
CC      acid sequence disclosed in GB-9928197.4, which is given in the present
CC      invention.
XX
SQ      Sequence 797 AA;

Query Match 5.5%; Score 259; DB 22; Length 797;
Best Local Similarity 20.5%; Pred. No. 3.2e-09;
Matches 191; Conservative 121; Mismatches 342; Indels 278; Gaps 40;
```

```

OY      116 TPLSLEELFAQESTEMGINFND-----YIPXYOGEOPNSEVVPPTLEPERKELLIRLY 169
Db      16 SPLALADFTIODIRVEGLQRTPESTVFENYLPVKGDITYN-----DTHGSAIKISLY 66
OY      170 ARLFENDGVNVP---RLKAAFYOSSGSGEFSALGSSHOKTEPPANITAALED--ITDESA 224
Db      67 ATGFEDDVRETDAGQLLTVIERPTIGSLNITGAKMLONDA---IKKNLESFGLAOSQY 123
OY      225 MD--LNGSIPRLRO-----TALYAARAVGYVIDLSIINSIGEVVYIIDL 269
Db      124 FNOATLQAVAGLKEEYLGKGLNIQTTPYTKLARNRVDITIIDGKSAKITIDIEPE- 182
OY      270 GEPYIDYRAVEVREGADDAKFTVADEVPPLIGDV-----FHGKYETKMLIE 320
Db      183 GNOYVSDRKLMR-----QMSLFEFGGIWTMLTRSGNQNEQKFPADMKEVT 226
OY      321 NASAEHGYFDGRWLDRSDVYLIPDNTADVSLITDTGYQYREDEVFTTIDPKTNQLTDP 380
Db      227 DFYQNNGYFDFRILIDTDIQTNEDKTKQTIKITVHEGGRFGKV---SIEGDINE----- 278
OY      381 DKLPYRKELLEOLLTVNMGSEAYNLOAVRALSNDLIATRFNMVMTETVPEREQIONOV 440
Db      279 ---VPRAELEKLLTMKPGKMYERQOMTAVLG-----EIQNRMG 313
OY      441 SFEQSSSRTEPAQVDESTEPEVIEVELTDGLMDISP-----IEFSASNLIODKL 492
Db      314 SAGAVYS-----EISVQPLNAETKIVDFVLHIEPGKIIYVNEIHITGNKKTKDE- 363
OY      493 NLVAAKARHLIYMPDD-----RVLAINHDDGVNRSILGRISDAVSAVARAILPDES 543
Db      364 -VVRRELQWESApydtskLQRSKERYELLGFPDNO-----FDAVPLAGTBDK- 411
OY      544 ENVEYIDPERTALANRKTPTADVQSKVPLXVFAASQKPRDQIGL--GMSDPTGLTYT 601
Db      412 ---VDL--NMSLTERST-----GSIDLSAGWYQDTG--LVM 440
OY      602 KFEHNLINRDYQAGAEARLSEDKKGKLYATKPLSHPLNDOLRATLGQOEVGHSTNG 661
Db      441 SAGVSDONLFGTGKSALRLASRSK--TTLNGSLSTPTDPTFADGVSLGY--DYGKAIFDP 496
OY      662 FDLSTRTLEHEISRSIIQNGGWR---TYSLRYLRLDKLKTQAPPTWDLR--VDPV 713
Db      497 RKASTSIQYKTTTA---GAGIRMSVPTVEYDRVNLGVAEHLTVTNKAPKHADYDFI 552
OY      714 N--GK-PSQEALLAGVAVHKTV-----ADNLVPMMGYQORYSLBEGSSGLVSDAMMAI 764
Db      553 KKYGKTDTGTSFRKGLYKGTGGRNKTDSALMPTRGY-----LTGVAEI 599
OY      765 ARAGISGVYSGDNAGSNRAHQMTG-----GIQAGYIWSDNFNHVPYR 808
Db      600 ALPGSKLQY-----YSATHNQTFPPFLSKFTTLMLGGEVGIAGY---GRTKEIPFF 648
OY      809 LRFEPAGDOSIRGYAHDSLPISDKGYLTGGQYLAVG---TAEYNYEFM-----KD 856
Db      649 ENFYGGGLGSVRYEGSGTIGP---KYDEYGEKISYGNKKANVASALLEPMPGAKDART 705
OY      857 LRLAVFGDIGNAYD-----KGFTNDTKIGAGVGVWAS 889
Db      706 VRLSLFLFADAGSVWDGKTYDDNSSSATGGRVQNTYGAGNTHKSTFTNRLRYSGAGAVTWLS 765
OY      890 PYGQVRVDVATGVKE---EGNPILKHFFIGTFP 919
Db      766 PLGPMKFSTAYPLKKRPEDEIQRFQGLGTTTF 797

RESULT 15
AAB23788
ID      AAB23788 standard; Protein; 797 AA.
XX
AC      AAB23788;
XX
DT      12-JAN-2001 (first entry)
```



PT New isolated outer membrane protein 85 of *Neisseria gonorrhoeae* and *N. meningitidis* useful for vaccine, therapeutic and diagnostic compositions for gonococcal or meningococcal infections -

Claim 41; Page 89-92; 98pp; English.

The present sequence represents an outer membrane protein (omp) 85 of *Neisseria meningitidis*. The omp polypeptides and polynucleotides are useful in compositions for use in the prevention, treatment and diagnosis of non-symptomatic gonococcal infection or meningococcal infection and symptomatic disease. They are also useful for the detection of hybridisation complexes. Antigens and antibodies specific omp proteins also provide diagnostic, therapeutic and prophylactic compositions for the treatment or prevention of the infections described above. The antibodies are useful for inducing a protective immune response in humans or animals with *N. gonorrhoeae*, *N. meningitidis*, or other *Neisseria* species. The proteins, antibodies and polynucleotide sequences of the present invention may also be used in the screening and development of chemical compounds such as drugs or vaccines.

**SQ** Sequence 797 AA;

Query Match 5.48; Score 253; DB 21; Length 797;

Best Local Similarity 20.4%; Pred. No. 8.4e-09;  
Matches 190; Conservative 121; Mismatches 343; Indels 278; Gaps 40;

QY	116	TPLELEEFADSETEMGINPMD-----PYEGOEOPNSEVVPPLPEBEKQELIKRL	169
Db	16	SPLAFAFETIODINVEGIQIRPEBTVNITLPAVKGDYTN-----DTHGSAITIKSLY	66
QY	170	ARLFNDGVNKP---RLKAFYSSOGSETSAIGSSHQKTEPYANIKAALED--ITQESA	224
Db	67	ATGFEDDVRVETADGOLLTIERPPTGISLINTAKMLONDA--IKNLSFGLAQSOY	123
QY	225	MD---LNGSIPRLNO-----TALYAARAVGYIIDLSITKNSIGEVDYIITHL	265
Db	124	FNQATLMOVAAGLKEEYLGRCKLNIOITPKYTKLARNRVIDITIDECKSAKITDIEE-	1822
QY	270	GEPLYIDYRAVEVNEGADDKAFPTVADEVPLIGDY-----FHAGKYETKMLIE	320
Db	183	GNQVYSDRKLAR-----QMSLTFGGIWTMLTRSNQNECKFPQDMKXYT	226
QY	321	NASAEHYCFQGRMLDREVDYILPNTADVLSILYDTQYRPDEYVFTIDPKTNQLTDP	360
Db	227	DFYONNGYFDRILDPDIQTNEDKTQOTIKITHEGGRFPMKGV--SIGDITNE----	278
QY	381	DKLPVKRELEOLLTVNMGEAYNLQAVRAALSNDLIATRYENMVTELVPEREODINOY	440
Db	279	---VPRAELEKLLTLMKPKKYEQOMTAVUG-----EIQNRMG	313
QY	441	SEEOSSSRPEPAOVDESTLEPIETVELDILMISP-----IEFSASNLIODKL	4922
Db	314	SAGYAVS-----EISVQPLPNAETKTVDFAVILHEPERKITYNEIHIITGNKTRDE-	363
QY	493	NLVAKARKHLHYDMPD-----RLAIINHODGVNRSLIGRSDAVSAVAARAILPDES	543
Db	364	-VVARERLOHESAPYDTSKLORSKERVELLGYFPNVV-----FDVAPLAGPDK-	411
QY	544	ENEVIDLPERITALANRKTTPADYVYOSKKVPLVYFVASDPRDQIGL--GWSQSTGRLYV	601
Db	412	---VDL--NMSLTERST-----GSULDSAGWQDGT--LYW	440
QY	602	KFEHNLINRQCYQAGALRLSEDKKGKYLATATPLSHPLMDOLARLQIOOEYFGHSYNG	661
Db	441	SAGVSONLFGTGSAAALRASRSK--TTLNGSLSFDPDYPETADGVSGCY--DYYGKAFLP	496
QY	662	FDLSTRLLEHISIKSIIONGWNV-----TYSRLRYLDKLTQAPPEYMDQD---VDVF	713
Db	497	RKASTSIKQYKTTTA---GAGIRMSVPYETBYDRVANGVLVAEHLVYNTYIKKAPKHADPE	552
QY	714	N--GK-PSOALLAGVAHVHTV-----ADNLVPMRGYRORYSLEVGSSGLVSDANMAI	764
Db	553	KKYGKTDGTOGSEFGKMWLYKGTGVGGRNKTDASALMPRTGCV-----LTGVNAEL	599

```

QY 765 ARAGISYSTRSGDNAYGSRNRHAKWTG-----GIQAYISINDFNHNHPYR 8008
Db 600 ALPGSKLOY-----YSATHNOTWPEPPLSKTPTLMLGGEVINGAGY---GRTKEI:PF 6488
QY 809 LRFPGAGDOSIRGYAHDSLPISDKGYLTGOVLAVG---TAEINYEFM-----KD 8566
Db 649 ENFYGGIGAGSYRGYESGTLGP---KYIDEYGEKISYSGNKKANSALLLEPPMGAKART 70505
QY 857 LRLAVFGDIGNAYD-----KGFYNDTKIGAGVYRMAS 889
Db 706 VRLSLEADAGSVMDQKTYDDNSSSATGGRVQNIYGAGNTHKSTPTNLRYSAGAGVWLS 7655
QY 890 PVGQVRYDVATGKVE--EGNFIKLHFFITGPF 919
Db 766 PLGPKKERYAYPLKKKPEDEIQRQFOLGTTGF 797

```

## RESULT 17

AA678603 standard; Protein; 896 AA.

AA AAG78603;  
AC

DT 20-NOV-2001 (first entry)

AA	DE	Lawsonia intracellularis protein SEQ ID NO: 5.
AA	DE	Lawsonia intracellularis protein SEQ ID NO: 5.

22 Htra; Pona; Hypc; Yefw; Abcl; Omp100; Lawsonia intracellularis infection;  
23 KW

KW  
XX

VACCINE.

03 LAWSONIA INTRACELLULARIS  
XX

UF 2001105101-A  
XX

FD 20 JUN - 2001  
XX

XX	2005/0750-300007	20007-100-02	XX
FF			FF

[illegible]

XX  
XX  
XX (FTE) FTEEN FTE INC.

XX	10/040700Z	1007	15M
----	------------	------	-----

PT prevent Lawsonia intracellularis infection -

PS Claim 12; Page 43-45; 67pp; Japanese.

23 The present invention provides isolated polynucleotides encoding HTRA  
22 CC  
21

intracellularly. The sequences can be used in vaccines for the prevention

CC of the invention.

Seq	Sequence	896 AA;
1	...	...
2	...	...
3	...	...
4	...	...
5	...	...
6	...	...
7	...	...
8	...	...
9	...	...
10	...	...
11	...	...
12	...	...
13	...	...
14	...	...
15	...	...
16	...	...
17	...	...
18	...	...
19	...	...
20	...	...
21	...	...
22	...	...
23	...	...
24	...	...
25	...	...
26	...	...
27	...	...
28	...	...
29	...	...
30	...	...
31	...	...
32	...	...
33	...	...
34	...	...
35	...	...
36	...	...
37	...	...
38	...	...
39	...	...
40	...	...
41	...	...
42	...	...
43	...	...
44	...	...
45	...	...
46	...	...
47	...	...
48	...	...
49	...	...
50	...	...
51	...	...
52	...	...
53	...	...
54	...	...
55	...	...
56	...	...
57	...	...
58	...	...
59	...	...
60	...	...
61	...	...
62	...	...
63	...	...
64	...	...
65	...	...
66	...	...
67	...	...
68	...	...
69	...	...
70	...	...
71	...	...
72	...	...
73	...	...
74	...	...
75	...	...
76	...	...
77	...	...
78	...	...
79	...	...
80	...	...
81	...	...
82	...	...
83	...	...
84	...	...
85	...	...
86	...	...
87	...	...
88	...	...
89	...	...
90	...	...
91	...	...
92	...	...
93	...	...
94	...	...
95	...	...
96	...	...
97	...	...
98	...	...
99	...	...
100	...	...

Query Match	5.08;	Score 238.5;	DB 22;	Length 896;
-------------	-------	--------------	--------	-------------

Best Local Similarity 20.0%, Pseud. NO. 9.0e-08;  
Matches 198; Conservative 137; Mismatches 391; Indels 239; Gaps 38;

QY 50 NQAKAGNPVLLTPEQIQARLNAGLNAK - PQSQALDVNFDDQSPIRIGEQSPPLGLD 108

Db 24 NAASKDPSIVLPEQINGSSNDELEQTLPMLLATAKN-----KG 65

QY 109 MSVIEETTPLSLEELFAQESTEMGINPNDYIPEYQGEQPNSEVVPPTLLEPEKPKLIKRL 1688

```

OY 226 DLNGSIPRLQRTALVAARAGVYDIDLSIRNSIGEVDIIH-----DLCEPVY 274
Db 156 --NGLIKK-NTIADYRIHGLKVLDPDVLITRLTIKGDHTDAKINAELKKIWEIG--YF 210
OY 275 IDYRA-VEVREGADDKAFTTVADVEPLLIGDVFHHGKYETKKNLIENASAEHGFDDRW 333
Db 211 SDVASISIESEGG---RLLVFTVQEKPKITDVVYQGSRAVSIDNLIAMSKKG----- 261
OY 334 LDRSV--DVILPNDTADVSLIYDTCTOYRFDEVFETIDPRTNOLITPPDKLPIVARELLE 391
Db 262 ---SVISORLLSODIQKTTDILYRKKEGYLAE--VAVETKEKENYSSA-----T 304
OY 392 QLTIVNNGEAVYNQAVRALSNDLIATRYFNMYNTEIYPER-----EOTIION 437
Db 305 LLTIVNECKKLYIKDVRIEGLEJTIKAK---TLKKEALATERNPLSWFTGTGLREYLER 361
OY 438 DOVFEEGSSSRTERPAQVDESTLEP-----VIEVELTDGILMDISPIEFSANLIQ 489
Db 362 DSIAI---SAYAMNHGYVDIQVASPEVTFNEKGIYITFRVKEGRYKIGKIDFGDILET 418
OY 490 DKLLVAAKARHLVMDPDDRYLAINHDDGVNRSLIGRISQAVSAVARAILPESENE--- 546
Db 419 NEQLKTKTKIDDKNRYEQYFSLYMODD--VKALTFEYSDYGAFAEVDL-ETTKNEEDA 475
OY 547 -----VIDLPERTAL-----ANKRTPADY----- 566
Db 476 TIDVFLIDKKQKVFRLRIIEGNTFRDNYILRELRLAOGDLFNGOHLRNSNCLNRLG 535
OY 567 -----QSKKVPYLYVPASDKPRDGOIGLWGSDTGTRLVTK---FEHNEINR 610
Db 536 YFNOVDTDLPTRKDDVDVLLVKV--QEARTGALTGGVGYSTHSGFVSGSISERNLMGK 593
OY 611 DGYOAGAEHLRLSEDKKGVLYATKPLSHPLNDQRLATLGVOQEFHGTNGFDL----- 664
Db 594 -GVILSEGTISKSSSLDLSTFNPRYDTD-----FGFSNNIYTLRLDEMMD 639
OY 665 -----STFLEHEISRSITONGMNRYSRLRYRLDKLTKQAPEWMDLPVDFVNGKP 717
Db 640 FRKTYGDTIRLFHPI-----GEYSSIFGVYRIDQYRLYDIPSTAPRSYLDY-QGKN 690
OY 718 SQALLAGAVAVHNTVADNLVNPWRGRQKRSLEVSGLVSDANMAMTARAGISGVYSGFD 777
Db 691 ISSVVGSGTFPDST--DSREPRSGKIAKLIVEYCGGGLGNDNPFKPIAELGCFYSI-- 746
OY 778 NAYGSNRAHQMTGGIOAGYIMSDNFHVPYRLRFFAGGQOSTGYAHDSLPSIDK-CYL 836
Db 747 ---SRKKNHIIHMKRTAGAAAYKNSKKPVVPDFPFITGGIDISIRGYDTEDLAPKDPFRGDE 803
OY 837 TCGOVLAVGTAENYEFMKDLRLAV--FGDIGNAYDKGFTND---TKIGAGVGVWASP 890
Db 804 IGGDRMAFLNLEIYMTFOPELGALVPFYDIDGIFQDTSVOTSNFSLKOSYGLLELWRSP 863
OY 891 VQOVR 895
Db 864 MGDLR 868

```

RESULT 18  
ABG17710  
ID ABG17710 standard; Protein; 474 AA.

XX ABG17710;

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #17701.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.  
XX

PN W0200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HXSE-) HXSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB; AAS81897.  
XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensic, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

XX Claim 20; SEQ ID NO 48069; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pat\_sequences.

XX Sequence 474 AA;

Query Match 5 0%; Score 235.5; DB 22; Length 474;

Best Local Similarity 45.2%; Pred. No. 6.5e-08;

Matches 56; Conservative 15; Mismatches 48; Indels 5; Gaps 2;

OY 795 GYIWSDNFNVHVPYRLRFFAGGQOSTGYAHDSLPSIDKGYLGGVAVGTAENYEFER 854  
Db 98 GWIETGDFPKVPPDLRFFAGGQOSTGYAHDSLPSIDKGYLGGVAVGTAENYEFER 157

OY 855 KDRLAVFGDIGNAVYDKGFTNDTKIGAGVGVWASPVGQVRVAVATGV--KEEGNFIKILH 912  
Db 158 GRWVGAVFVDSGEAVSDIRSDFKTGTGVGVWESPVGPIKIDFANVADKDEHG---LQ 214

OY 913 FFIIG 916  
Db 215 FYIIG 218

RESULT 19  
AAV34506  
ID AAV34506 standard; Protein; 775 AA.

XX AAV34506;

XX 25-AUG-1999 (first entry)

XX Porphyromonas gingivalis protein PG45.  
XX



```
XX Porphyromonas gingivalis; PG: periodontal disease; gingivitis;
KM vaccine; antigenic.
XX
XX Porphyromonas gingivalis.
XX
XX MO929870-A1.
XX
XX 17-JUN-1999.
XX
XX 10-DEC-1998: 98MO-AU01023.
XX
XX 04-AUG-1998: 98AU-0005028.
XX 10-DEC-1997: 97AU-0000839.
XX 31-DEC-1997: 97AU-0001182.
XX 30-JAN-1998: 98AU-0001546.
XX 10-MAR-1998: 98AU-0002264.
XX 09-APR-1998: 98AU-0002911.
XX 23-APR-1998: 98AU-0003128.
XX 05-MAY-1998: 98AU-0003338.
XX 22-MAY-1998: 98AU-0003654.
XX 29-JUL-1998: 98AU-0004917.
XX
XX (CSLC-) CSL LTD.
XX
XX Agius CT, Barr IG, Hocking DM, Margetis MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
XX MPI: 1999-385613/32.
XX N-PSDB; AAX91724.
XX
XX Antigenic Porphyromonas gingivalis peptides for preventing
XX gingivitis
XX
XX Claim 1, Page 487-489; 588pp; English.
XX
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX AAY34583. AAX91802 to AAX91889 represent PCR primers used in the
XX isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX activity with a vaccine mechanism of action. The PG polypeptides can be
XX used as vaccines especially against Porphyromonas gingivalis. Probes can
XX be used to detect Porphyromonas gingivalis in standard hybridisation
XX assays. Porphyromonas gingivalis is involved in periodontal disease
XX especially gingivitis.
XX
XX Sequence 775 AA:
SQ
Query Match 4.7%; Score 224.5; DB 20; Length 775;
Best Local Similarity 20.6%; Pred. No. 7.4e-07;
Matches 155; Conservative 90; Mismatches 290; Indels 217; Gaps 34;
QY 300 PLLIGOVFNHGKYEETKKNLEENASAEHGFDDGWLDRSDVILPDN-VADVSLIYDTGTQ 358
DB 109 PFVFIQV-----SDSRAKATNLRHGXFDAK-VKSVYTTLKDKLAKISYVDWASP 163
QY 359 YRFDEVEFTIDPKTNQLTDPDKLPVKRELLQLLTVNMGEAYNL-----QAVRALS 411
DB 164 YHVDSTIRPI-----STFSDILAYKOTPS---LIRKGDQFNLAKEHERQTISALL 213
QY 412 NDLIATRFNMVNTLIVFEREQIONDOVSFEQSSSSRTEPA-----QVDESTL----- 460
DB 214 RD---NGYFRPRQDIYBADTLVVGAVCLRAKLSSEDPQAMRPWRIGKRAVLGGMN 270
QY 461 -EPVIEIVELTD-GIL---MDISP-----IEFSASNLIDKLNIVAAKARHLVMPDD 508
DB 271 GSPSDSLEVEEMKALVYRKMEVPRKILAKRRFFSGNLYRQK----- 313
QY 509 RVLAINHDDGVRSILIGRISDAVSAVARAILPDESENEVIDLPERTAL-----ANKTPA 563
DB 314 -----DDETRKSLARIG-AFSYIDLNFLORDSISGLLDVRLTLTLDKPMDALET-- 363
QY 564 DVGOSK-----VPLYVFASDKPRDGOIGLGM---GS---DTGTRLVYKFEHNLINRDG 612
```

```
DB 364 -LFTSKSNDFIGPGLNFALARNVFGGGENLSWMNIGSSYEMETGNR-PENSSNRLIDINS 421
QY 613 YQAGAEIRLSEDKKGVKIYATKPLSHPIINDOLRAT----- 647
DB 422 YNMNTAVNLSPFSIVFPGLDKYRYYPPTTFQASATALNBAHFMSYSPGFSTYERQP 481
QY 648 -----LGYOEVGHSTNGFDLSTRT---LEHISRSIIONGGMNNTISLRFLD 694
DB 482 SKERHAIPLKLYNLGLHOTETFOAITANNPPLLSLOSOFLAQMGIYTFN-----K 536
QY 695 KLKTOAPPETWQDLPVDFVNGKPSQEAALLAGVAHVKTADMLVN-PMGKY-----ROBS 748
DB 537 SVSEKSPHILMQFGLSAGMLNLITYLAACKKXSDT--KNFVGVPFQSFKAIGELRYS 594
QY 749 LEVGSSGLVSDANMAIA-RAGISGVSEGDNAYSNRAHOMTGTGIACYIWSDFNHVY 807
DB 595 YTI-----DRNOSLATRFGTGVYISYG-----NMRAVAPY 623
QY 808 RLRFAGGDOSIRGYAHDSL-----SPISDKGYL---TGQVLAVGTRAEYEFMKDLRL 859
DB 624 SEQFYVGANSIRAFYVRSIGPRFNPDSQYSLDOVGFEKLEANVEYRGKLFGLDHA 683
QY 860 AVFGDIGNAY-----DKGFTNDTRIGAGVGRMASPVQGVREVDVATG 901
DB 684 AVFLDAGNWNLLRSDSSRPGALSEVGSYSNFNLMSIALGTGVGLRIDLAVLRVDVFG 743
QY 902 VK-----EEG-----NPIKLHFEIGTPEF 919
DB 744 LHLPEYNTGKKGYNIIPRFKALIGFHLAVGYPF 775
RESULT 20
AAY34379
ID AAY34379 standard; Protein: 803 AA.
XX
XX AAY34379:
XX
XX 25-AUG-1999 (first entry)
XX
XX Porphyromonas gingivalis protein PG45.
XX
XX Porphyromonas gingivalis; PG: periodontal disease; gingivitis;
KM vaccine; antigenic.
XX
XX Porphyromonas gingivalis.
XX
XX MO929870-A1.
XX
XX 17-JUN-1999.
XX
XX 10-DEC-1998: 98MO-AU01023.
XX
XX 04-AUG-1998: 98AU-0005028.
XX 10-DEC-1997: 97AU-0000839.
XX 31-DEC-1997: 97AU-0001182.
XX 30-JAN-1998: 98AU-0001546.
XX 10-MAR-1998: 98AU-0002264.
XX 09-APR-1998: 98AU-0002911.
XX 23-APR-1998: 98AU-0003128.
XX 05-MAY-1998: 98AU-0003338.
XX 22-MAY-1998: 98AU-0003654.
XX 29-JUL-1998: 98AU-0004917.
XX
XX (CSLC-) CSL LTD.
XX
XX Agius CT, Barr IG, Hocking DM, Margetis MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
XX MPI: 1999-385613/32.
XX N-PSDB; AAX91597.
XX
XX Antigenic Porphyromonas gingivalis peptides for preventing
```

```

PT gInglvltis
XX Claim 1; Page 343-344; 588bp; English.
XX
XX AA91536 to AA91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY4318 to
CC AAY4583. AA91802 to AA91989 represent PCR primers used in the
CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The Pg polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gInglvltis.
XX
SQ Sequence 803 AA:
Query Match 4.7%; Score 224.5; DB 20; Length 803;
Best Local Similarity 20.6%; Pred. No. 7.8e-07;
Matches 155; Conservative 90; Mismatches 290; Indels 217; Gaps 34;
OY 300 PLLIDGVFHGKYEETKKNLEENASAEHGFDCGMDRSDVILPDN-TADVSLITYDTGQ 358
DB 137 PVFISQV-----KSDSRKAKVATNLRHGYFDAK-VKSVYTTTKKSLAKISYVDMASP 191
OY 359 YRFDEVVFETIDPKTNQLTDPDKLPVKRELLQLLTVMGEAVNL-----QAVRAIS 411
DB 192 YHDSITLPLT-----STFPDSILAYRQTPS---LIKKDQFNLAKLHEERQTSALL 241
OY 412 NDLITRFFNMVNTVEIPPEREQIONDOVSFEQSSSRTEPA-----QVDESTL----- 460
DB 242 RD---NGYVFRPDIIYEADTLVAGAVCLRAKLESDPPQAMRPMRIGKRTAVLGMN 298
OY 461 -EPVLETVELTD-GIL---MDISP-----IEFSASNLIDOKNLVAKAKRLHMDPD 508
DB 299 GSPSPDSLEVEDMKVLYRKMPVPRITAKRFRFSGNLYRK----- 341
OY 509 RVLAINHDDGVNRSLIGRISDAVSAVARAILPDESENEVIDPERPAL-----ANKRTPA 563
DB 342 -----DDETRKRLARIG-AFVYIDLNFLORDSISGLDVRLLTLTLKPRDASLET--- 391
OY 564 DVGYSKK-----VPLVYVVASDKPRDQIGLGM---GS---DTGRLVYKFEHNLINRDG 612
DB 392 -LFTSKSNDFTGPGINFAFALARRNVEGCGENLSMIGNISYEMETGNR-PENSSNRLIDNS 449
OY 613 YQAGELRLSEDDKGVKVLATPKRLSHPLNDOLAR----- 647
DB 450 YNMNTAVNLSRPSIVFPGCLLDKYYVYPTTTTFQASATALNRAHYFSMSEGEFTVEEP 509
OY 648 -----LQYQOEVEFGHSTNGFDLSTRT---LEHEISRSIIIONGGMRTYSRLRYLD 694
DB 510 SKENRHATFPLKLANVNLGHOTETFOAITANNRPLLISQSFLOMKGITYTFN-----K 564
OY 695 KLKQVAPRETQDLPVDVNVNGKPSQEBALLAGVAVHKYVADNLVN-PMRGY-----RQRY 748
DB 565 SVSEKSPHNLMMQFELSAGNLLNLIYLAAGKKYSDT---KNVGVFESQFIATGELRYS 622
OY 749 LEVGSGLVSDANMAIA-RAGISGVYSPFGDNAGYSGNRAHMQMGIOAGIWMNDNHPY 807
DB 623 YTI-----DRNOSLARFRGTGVITYTG-----NMRAVAY 651
OY 808 RLRFAGGDSIRGYANDSL-----SPISDKGYL---TGGOVLAVGTAENYEFKKDLRL 859
DB 652 SEQFYVGANSIRAFYVASTIGRGRENPDSDQSYLDQVGEKLEANEYRCKLEFGDLHA 711
OY 860 AVFDGIGNAV-----DKGFTNDTKIGAGVGVWASPVGQVRVAVATG 901
DB 712 AVFLDAGNVMVLLRBDSSRRGALSEVGSVSNFLNSIALGTGVLGYDLAFLVAVDVGFG 771
OY 902 VK-----EEG-----NPIKILHFIQTPE 919
DB 772 LHLPTNYGKKGYNIIPRKDAIGRHAVGTFP 803

```

```

RESULT 21
AA74862
ID AAY74862 standard; Protein; 615 AA.
XX
XX AA74862;
AC
XX 21-MAR-2000 (first entry)
DT
XX Neisseria meningitidis ORF 286 protein sequence SEQ ID NO:1198.
DE
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
XX Neisseria meningitidis.
OS
XX
XX MO9957280-A2.
PN
XX 11-NOV-1999.
PD
XX
XX 30-APR-1999; 99WO-US09346.
PF
XX
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
PI
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzo M, Rappoli R, Ratti G, Scalzo E, Scarselli M;
PI Tettein H, Venter JC;
XX
DR WPI: 2000-062150/05.
DR N-PSDB; AA253624.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 669; 1453bp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254773 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 615 AA:
Query Match 4.7%; Score 220; DB 21; Length 615;
Best Local Similarity 22.3%; Pred. No. 1.1e-06;
Matches 131; Conservative 72; Mismatches 236; Indels 148; Gaps 19;
OY 462 PVIEFVELTDGILMDISPIEFASNLIDOKNLVAKAKRLHMDPDRLALNHDDGVNR 521
DB 48 POTESVKLKKPRPVRIIDQDSIKDVEERHPLITQOQEEVLDKQGTGLAEBADPNVKT 107
OY 522 SI--IGRISDAVSA-----VARAILPDESENEVIDPERPALAN 558
DB 108 MLRSKGYFSSKKSLETKGQATYVHTTGPRTKIANVGAVALIDILSDGSLAETRYNALEN 167

```

```

QY 559 RKTPA-----DYQSK-----VPLYVFASDKP 582
      : : : :
      : : : :
Db 168 WQPVGSDFDODSMENKSTVIGAVTRKAYPLAKGNTRAAVNPDATADLVVDSGRP 227
      : : : :
      : : : :
QY 583 RDQIGLGWGSDTGTR-----LYTKFEHNLINRDGYAGAEI- 619
      : : : :
      : : : :
Db 228 ----IAFGDFEITGTQRYPEQIVSGLARFQPGTPYDLDDLDFQOAL-EQNGHYSGASVQ 282
      : : : :
      : : : :
QY 620 ----RLSEDKKGYLKYATKPLSHPLNDQLRAT-----LG-----YQGEVFGHSTNG 661
      : : : :
      : : : :
Db 283 ADFDRIGQDRAVPKAVSYTEVKRHKLETGIRLDSEYGLGKIAIYYNLFNKGYIGSVYMD 342
      : : : :
      : : : :
QY 662 FDLSTRTLEHRSITIQNGW---NRTYSLRYRLDKLKTQAPPETMODLPVDFVNGKPS 718
      : : : :
      : : : :
Db 343 MDEYETTLAAGISQPRMYRCNWTMTSNVSYNSTQNLKRAFSGCIWVVRACIDARLG 402
      : : : :
      : : : :
QY 719 QEALLAGVAV-----HKT-----VADNLVNPARGYRQRYSLSEVSSGLVSD 759
      : : : :
      : : : :
Db 403 AEFLAEGRKIPGSDIDIGNSHATMLTASMKRQLNNVLHPENGHYLDGKIGTTLGAFIJS 462
      : : : :
      : : : :
QY 760 ANN-ATAAGISGVYSFGDNAYGSNRAHOMTGTI---QAGY-IMSDNENHVPYRLREF 812
      : : : :
      : : : :
Db 463 TALIRTSARAG-----YFTPEKKLGTFTIRGQAGTYVARDNAN-VPSGLMFR 510
      : : : :
      : : : :
QY 813 AGGDQSTIRGYAHDLSPISDKGYLTGGQVLAAGTAENYEFMKDLRLAVFGDIGNAYDKG 872
      : : : :
      : : : :
Db 511 SGGASSVRYGELDISIGLAPNGSVLPERALLVGSLEQVLPFTRTLSCAVFHDMDGDAANF 570
      : : : :
      : : : :
QY 873 FTMDTIRIGAGVGRMASPVQGVRYDVATGVKEEGNPILKHFFIGTPE 919
      : : : :
      : : : :
Db 571 KRMKLHSGSLGVWFSPLAFPSFDIAYGHSDK--KIRMHISLGTRF 615
      : : : :
      : : : :

```

## RESULT 22

AA99623 standard; Protein; 609 AA.

AA99623:

03-JAN-2001 (first entry)

Neisseria meningitidis BASB040 putative protein sequence #1.

BASB040: bacterial disease; respiratory tract infection; bacteraemia;

meningitis; cancer; autoimmune disease.

Neisseria meningitidis.

WO200034480-A1.

15-JUN-2000.

02-DEC-1999; 99WO-EP09560.

07-DEC-1998; 98GB-0026886.

(SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;

WPI: 2000-423426/36.

N-PSDB; AAA48507.

Novel BASB040 polypeptides of Neisseria meningitidis useful for

diagnostic, prophylactic and therapeutic purposes against microbial

diseases comprise a specific amino acid sequence

Claim 4; Page 59-60; 98pp; English.

The present sequence is a putative version of the Neisseria

meningitidis strain ATCC13090 BASB040 protein sequence. This protein is

similar to the D15 outer membrane protein of the bacterium. The protein.

its gene, antibodies, antagonists and agonists can be used to diagnose CC and treat bacterial diseases such as those leading to upper respiratory tract infections, bacteraemia and meningitis. In addition, they can be used in vaccines for use against cancer and autoimmune diseases.

Sequence 609 AA;

Query Match 4.6%; Score 219; DB 21; Length 609;

Best Local Similarity 22.0%; Pred. No. 1.3e-06;

Matches 130; Conservative 74; Mismatches 232; Indels 154; Gaps 18;

```

QY 462 PVIEVELTDLGILMDISPIEFSSASNLIQDKNLVAAKRHLIYDMPDRLAINHDDGYNR 521
      : : : :
      : : : :
Db 42 PDTESVRLKPFPRIPRIDQSEIKDWEHLPLITQOGEVLDEKQGFLEAPADNKT 101
      : : : :
      : : : :
QY 522 SI--LGRISDAVSA-----VAAIIPDESENVYIDLPELTALAN 558
      : : : :
      : : : :
Db 102 MLRSKGYFSSKVSLETKDGAYTVHTPGPRRKIANVGAIIIGDILSDGNIAEYRNALEN 161
      : : : :
      : : : :
QY 559 RKTPA-----DYQSK-----VPLYVFASDKP 582
      : : : :
      : : : :
Db 162 WQPVGSDFDODSMENKSTVIGAVTRKAYPLAKGNTRAAVNPDATADLVVDSGRP 221
      : : : :
      : : : :
QY 583 RDQIGLGWGSDTGTR-----LYTKFEHNLINRDGYAGAEI- 619
      : : : :
      : : : :
Db 222 ----IAFGDFEITGTQRYPEQIVSGLARFQPGTPYDLDDLDFQOAL-EQNGHYSGASVQ 276
      : : : :
      : : : :
QY 620 ----RLSEDKKGYLKYATKPLSHPLNDQLRAT-----LG-----YQGEVFGHSTNG 661
      : : : :
      : : : :
Db 277 ADFDRIGQDRAVPKAVSYTEVKRHKLETGIRLDSEYGLGKIAIYYNLFNKGYIGSVYMD 336
      : : : :
      : : : :
QY 662 FDLSTRTLEHRSITIQNGW---NRTYSLRYRLDKLKTQAPPETMODLPVDFVNGKPS 718
      : : : :
      : : : :
Db 337 MDEYETTLAAGISQPRMYRCNWTMTSNVSYNSTQNLKRAFSGCIWVVRACIDARLG 396
      : : : :
      : : : :
QY 719 QEALLAGVAV-----HKT-----VADNLVNPARGY-----RQRYSLSEVGS 753
      : : : :
      : : : :
Db 397 AEFLAEGRKIPGSDIDIGNSHATMLTASMKRQLNNVLHPENGHYLDGKIGTTLGFIJS 456
      : : : :
      : : : :
QY 754 SGLVSDANMAIARAAGISGVYSFGDNAYGSNRAHOMTGTI---QAGYIMSDNENHVPYRL 809
      : : : :
      : : : :
Db 457 TALIRTS---ARAG-----YFTPEKKLGTFTIRGQAGTYVARDNADVPISGL 501
      : : : :
      : : : :
QY 810 RFFAGDQSTIRGYAHDLSPISDKGYLTGGQVLAAGTAENYEFMKDLRLAVFGDIGNAY 869
      : : : :
      : : : :
Db 502 MFRSGGASSVRYGELDISIGLAPNGSVLPERALLVGSLEQVLPFTRTLSCAVFHDMDGDA 561
      : : : :
      : : : :
QY 870 DKGFTNDTKIGAGVGRMASPVQGVRYDVATGVKEEGNPILKHFFIGTPE 919
      : : : :
      : : : :
Db 562 ANFRMKLKHGSLGVWFSPLAFPSFDIAYGHSDK--KIRMHISLGTRF 609
      : : : :
      : : : :

```

## RESULT 23

AA99624 standard; Protein; 609 AA.

AA99624:

03-JAN-2001 (first entry)

Neisseria meningitidis BASB040 putative protein sequence #2.

BASB040: bacterial disease; respiratory tract infection; bacteraemia;

meningitis; cancer; autoimmune disease.

Neisseria meningitidis.

WO200034480-A1.

15-JUN-2000.

02-DEC-1999; 99WO-EP09560.

```

PR 07-DEC-1998: 98GB-0026886.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Ruelle J:
XX
XX WPI: 2000-423426/36.
XX N-PSDB: AAA48508.
XX
PT Novel BASB040 polypeptides of Neisseria meningitidis useful for
PT diagnostic, prophylactic and therapeutic purposes against microbial
PT diseases comprise a specific amino acid sequence
XX
XX Claim 4: Page 60-61; 98pp; English.
XX
XX The present sequence is a putative version of the Neisseria
XX meningitidis strain ATCC13090 BASB040 protein sequence. This protein is
XX similar to the D15 outer membrane protein of the bacterium. The protein,
XX its gene, antibodies, antagonists and agonists can be used to diagnose
XX and treat bacterial diseases such as those leading to upper respiratory
XX tract infections, bacteraemia and meningitis. In addition, they can be
XX used in vaccines for use against cancer and autoimmune diseases.
XX
SQ Sequence 609 AA:
Query Match 4.6%; Score 219; DB 21; Length 609;
Best Local Similarity 22.0%; Pred. No. 1.3e-06;
Matches 130; Conservative 74; Mismatches 232; Indels 154; Gaps 18;
QY 462 PVEYVELDGLIMDISPIEFSASNLIQDKLNLVAAKARHLVMPDDRVLAINHDDGVNR 521
DB 42 PDIESVYKLRKPFVRLIDTQDSEIKDWEEHLPLITQOQEEVLDEQGTGFLAEAPDVKVT 101
QY 522 SI--LGRISDAVSA-----VARATLPRESENEVIDLEPRALAN 558
DB 102 MLRSKGYFSSKYSLTEKDGAYVYHITPGPRTKIANGVALLGILSDGNIAEYRNALEN 161
QY 559 RKTPTA-----DYYQSKK-----VPLYVFVASDKP 582
DB 162 WOOPVSGDFDDQDQSWENSKYSVLGAVTRKAYPLAKLGNTAAVNPDTATADLVNVDVSGRP 221
QY 583 RQOQIGLGMGSDTGR-----LVTKFEHNLINRDGYOAGAEI- 619
DB 222 ----IAFGDFEITGTQRYPEQIVSGIARFQPGMPYDLDDLDFQOAL-BQNGHYSGASVQ 276
QY 620 ----RLSEDKKGVKLYATKPLSHPLNDOLRAT-----LG-----YQOEYFGHSTNG 276
DB 277 ADPDRLOGDVRVVKSVTVKRRHKLETGIRLDESEGLGKIAVDYNNLKNKGYIGSVWD 336
QY 662 FDLSTRITLHEHSRIIONGWM---NRTYSLRYRLDKLTQAPRPETWODLPDVFVNGKFS 718
DB 337 MDKYETTLAAGISQPRNRYNGNWTNSYNRSTQNLERKRAFSGGIWMYVRADRAGIDARLG 396
QY 719 QEALLAGVAV-----HKT-----VADNLVNPMRGY-----RQRYSLVSGS 753
DB 397 AEFLEGRKIPGSDIDLGNHATMLTASWKRLQNLNVLHPNGHYLDGKITGLTGLFELS 456
QY 754 SGLVSDAMKATIRACISGVYVSFGDNAYGSRNRAHQMTGT-----QAGYIWSDFNHHVYRL 809
DB 457 TALLRTS---ARAG-----YFETPENKKLGTFIIRGQAGYVARDNADVPSSL 501
QY 810 RFFAGGDSIRGYAHDSLPSIDKGYLGGQVLAAGTAEVNNEFKMDRLAVFGDIGNAV 869
DB 502 MFRSGASSVNGYEIDSLGACPGNSVLERPALLVGSLSELEYLPFTRTLISGAVFHMDGMA 561
QY 870 DKGFTNDTKIAGVGVNMAPYGVQVRVDVATGVKKEGNPIKLFHFIIGTPE 919
DB 562 ANFKRMKLRKHSGLGVNMFSPFLAPSPFDIAYGHSDK--KIRHISLGTRF 609
RESULT 24
AAV99625
ID AAV99625 standard; Protein: 587 AA.

```

```

XX
XX AAV99625;
XX
XX 03-JAN-2001 (first entry)
XX
XX Neisseria meningitidis BASB040 putative protein sequence.
XX
XX DE
XX
XX KW BASB040; bacterial disease; respiratory tract infection; bacteraemia;
XX meningitis; cancer; autoimmune disease.
XX
XX OS Neisseria meningitidis.
XX
XX PN WO200034480-A1.
XX
XX PD 15-JUN-2000.
XX
XX PF 02-DEC-1999; 99WO-EP09560.
XX
XX PR 07-DEC-1998; 98GB-0026886.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX PI Ruelle J:
XX
XX WPI: 2000-423426/36.
XX N-PSDB: AAA48509.
XX
PT Novel BASB040 polypeptides of Neisseria meningitidis useful for
PT diagnostic, prophylactic and therapeutic purposes against microbial
PT diseases comprise a specific amino acid sequence
XX
XX Claim 4: Page 62; 98pp; English.
XX
XX The present sequence is a putative version of the Neisseria
XX meningitidis strain H44/76 BASB040 protein sequence. This protein is
XX similar to the D15 outer membrane protein of the bacterium. The protein,
XX its gene, antibodies, antagonists and agonists can be used to diagnose
XX and treat bacterial diseases such as those leading to upper respiratory
XX tract infections, bacteraemia and meningitis. In addition, they can be
XX used in vaccines for use against cancer and autoimmune diseases.
XX
SQ Sequence 587 AA:
Query Match 4.5%; Score 215; DB 21; Length 587;
Best Local Similarity 22.1%; Pred. No. 2.3e-06;
Matches 132; Conservative 73; Mismatches 225; Indels 166; Gaps 20;
QY 462 PVEYVELDGLIMDISPIEFSASNLIQDKLNLVAAKARHLVMPDDRVLAINHDDGVNR 521
DB 20 PDIESVYKLRKPFVRLIDTQDSEIKDWEEHLPLITQOQEEVLDEQGTGFLAEAPDVKVT 79
QY 522 SI--LGRISDAVSA-----VARATLPRESENEVIDLEPRALAN 558
DB 80 MLRSKGYFSSKYSLTEKDGAYVYHITPGPRTKIANGVALLGILSDGNIAEYRNALEN 139
QY 559 RKTPTA-----DYYQSKK-----VPLYVFVASDKP 582
DB 140 WOOPVSGDFDDQDQSWENSKYSVLGAVTRKAYPLAKLGNTQAAVNPDTATADLVNVDVSGRP 199
QY 583 RQOQIGLGMGSDTGR-----LVTKFEHNLINRDGYOAGAEI- 619
DB 200 ----IAFGDFEITGTQRYPEQIVSGIARFQPGMPYDLDDLDFQOAL-BQNGHYSGASVQ 254
QY 620 ----RLSEDKKGVKLYATKPLSHPLNDOLRAT-----LG-----YQOEYFGHSTNG 254
DB 255 ADPDRLOGDVRVVKSVTVKRRHKLETGIRLDESEGLGKIAVDYNNLKNKGYIGSVWD 314
QY 662 FDLSTRITLHEHSRIIONGWM---NRTYSLRYRLDKLTQAPRPETWODLPDVFVNGKFS 718
DB 315 MDKYETTLAAGISQPRNRYNGNWTNSYNRSTQNLERKRAFSGGVW-----YVADRAG 368
QY 719 QEALLA-----GVAV-----HKT-----VADNLVNPMRGY-----RQRY 747

```

Db 369 IDARLGAFLAEGRRKIPGSADVLDGNSHATMTLASMCKROLLNNVLPENGLDCKIGTTL 428  
QY 748 SLEVGSSGLVSDANMAIARAGISGVSPGDNAYSNRAHQMTGFI---QAGYIMSDNFN 803  
Db 429 GTFLSSTALIRTS---ARAG-----YFTPEPKKLGFTFIIRGAGGYVARNA 473  
QY 804 HVPYRLRFPAGDGSIRGYAHDSLSPISDKGYLTGGQVLAAGTAEYNEFKDLRLAVFG 863  
Db 474 DVPSGLMFRSGGASSVGYELDSIGLAGPNGSVLPERALLVSGLSLEYQLPFTRTLGAVERH 533  
QY 864 DIGNAYDKGFTNDKIGAGVGRWASPVGVAVDVAAGVKEGNPIKLHFIPTGPF 919  
Db 534 DMGDAANFRMKLKHSGSLGVRFWFSPLAPPSFDIAYGHSK--KIRWHISLGTFR 587  
RESULT 25  
AAY74861  
ID AAY74861 standard; Protein: 615 AA.  
XX AAY74861;  
AC  
XX 21-MAR-2000 (first entry)  
XX  
DE Neisseria meningitidis ORF 286 protein sequence SEQ ID NO:1196.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
XX antibacterial; gene therapy.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO9957280-A2.  
XX  
XX 11-NOV-1999.  
PD  
XX 30-APR-1999; 99W0-US09346.  
PF  
XX 01-MAY-1998; 98US-0083758.  
PR 31-JUL-1998; 98US-0094869.  
PR 02-SEP-1998; 98US-0098994.  
PR 02-SEP-1998; 98US-0099062.  
PR 09-OCT-1998; 98US-0103749.  
PR 09-OCT-1998; 98US-0103794.  
PR 09-OCT-1998; 98US-0103796.  
PR 25-FEB-1999; 99US-0121528.  
XX  
XX (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
PI Tettelin H, Venter JC;  
XX  
DR MPI: 2000-062150/05.  
DR N-PSDB; AA253623.  
XX  
PT Novel Neisserial polypeptides predicted to be useful antigens for  
PT vaccines and diagnostics -  
XX  
PS Claim 2: Page 667; 1453pp; English.  
XX  
CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941  
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides  
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC presence of Neisseria bacteria (e.g. meningitis and septicemia), to detect the  
CC presence of Neisseria bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.  
XX  
SQ Sequence 615 AA;  
Query Match 4.5%; Score 215; DB 21; Length 615;  
Best Local Similarity 22.18; Pred. No. 2,4e-06;  
Matches 132; Conservative 73; Mismatches 225; Indels 166; Gaps 20;  
QY 462 PVIEVELTDCILMDIPSEFASNLIDCKLNLVAKARHLYDMPDDRVLAINHDDGYNR 521  
Db 48 PDTESVKLRKPFVILIDTQDSEIKDMVEHLPLITQDQEEVLDEKQTFLEAPADNKT 107  
QY 522 SI--LGRISDAVA-----VARAILPDESENEVIDLPERTALAN 558  
Db 108 MLRSKGYFSSKVSLETKDGAVTHITPGPRKIANVGAIIIGDILSDCNLEAYRNALEN 167  
QY 559 RKTPA-----DYQSKR-----VPLVYVASKP 582  
Db 168 WQDPVSDPDQDSEWNSKTSVLGAVTRKAYPLAKLGNTOAAVNPDTATADLNVDVDSGRP 227  
QY 583 RDGQIGLGWGSDTGTR-----LYTKFEHNLINRDGYOGAEL- 619  
Db 228 ----IARGDEFITGTQRPEDIVSGLARFQGMPIYDLDLDFQAL-EQNGHTSGASVQ 282  
QY 620 ----RLSEDKKGVKLYATKPLSHPLNDQLRAT---LG-----YQGEVEGSHSTNG 661  
Db 283 ADPFRLGDDRVPVAVSVTEVKRHKLENGIRLDSEYGLGKLAIDYVNLFNKGYIGSVVWD 342  
QY 662 FDLSTRTLEHESISIIONGWM---NRTYSLRRLDKLKTQAPPETWODLPVDFVNGKPS 718  
Db 343 MDKYETTLAAGISQPRNRYGNYWTSNVSYNSTQNLKRAFGSGVW-----YVRDRAG 396  
QY 719 QEALLA-----GVAY-----HKT-----VADNLVNPNGY-----RGRY 747  
Db 397 IDARLGAFLAEGRRKIPGSADVLDGNSHATMTLASMCKROLLNNVLPENGLDCKIGTTL 456  
QY 748 SLEVGSSGLVSDANMAIARAGISGVSPGDNAYSNRAHQMTGFI---QAGYIMSDNFN 803  
Db 457 GTFLSSTALIRTS---ARAG-----YFTPEPKKLGFTFIIRGAGGYVARNA 501  
QY 804 HVPYRLRFPAGDGSIRGYAHDSLSPISDKGYLTGGQVLAAGTAEYNEFKDLRLAVFG 863  
Db 502 DVPSGLMFRSGGASSVGYELDSIGLAGPNGSVLPERALLVSGLSLEYQLPFTRTLGAVERH 561  
QY 864 DIGNAYDKGFTNDKIGAGVGRWASPVGVAVDVAAGVKEGNPIKLHFIPTGPF 919  
Db 562 DMGDAANFRMKLKHSGSLGVRFWFSPLAPPSFDIAYGHSK--KIRWHISLGTFR 615  
RESULT 26  
AAB46311  
ID AAB46311 standard; Protein: 916 AA.  
XX  
XX AAB46311;  
AC  
XX 05-APR-2001 (first entry)  
DT  
XX H. pylori HPS120 protein.  
DE  
XX Microbial infection; antibacterial; Helicobacter pylori infection;  
KW vaccine; screening.  
XX  
OS Helicobacter pylori.  
XX  
PN WO200073502-A2.  
XX  
PD 07-DEC-2000.  
XX  
PF 31-MAY-2000; 2000W0-EP05024.  
XX  
XX 31-MAY-1999; 99DE-1024965.  
PR 17-JUN-1999; 99DE-1027740.  
PR 21-JUL-1999; 99DE-1034029.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 PA (CREA-) CREATOGEN GMBH.  
 XX  
 PI Apfel H, Fuchs TM, Gibbs CP, Hueck CJ, Meyer TF;  
 DR WPI; 2001-049948/06.  
 DR N-PSDB; AAF25588.  
 XX  
 PT Preparing an agent for diagnosis or control of microbial infection,  
 PT useful particularly against Helicobacter, based on identification of  
 XX essential genes in defective mutants.  
 XX  
 PS Claim 37; Page 236-239; 366pp; German.  
 XX  
 CC This invention describes a novel preparation of an agent (A) for  
 CC detection, prevention and/or treatment of microbial infection by:  
 CC (i) identifying essential genes (I) and corresponding polypeptides  
 CC (II); (ii) identifying compounds that are directed against (II) and  
 CC inactivate the microbe; (iii) testing these for suitability for use; and  
 CC (iv) formulating selected (A). Identifying essential genes (I) comprises  
 CC preparation of gene-deficient microorganisms by conditional antisense  
 CC inhibition (CAI) and/or subtractive recombination mutagenesis (SRM),  
 CC then determining viability and/or survival of the deficient organisms.  
 CC The products of the invention have antibacterial activity. (A) (which may  
 CC be a nucleic acid (Ia), vector or host cell containing (Ia), derived  
 CC polypeptide (IIa), or fragments, (IIa)-specific antibodies or their  
 CC fragments or an inhibitor of (IIa)) are particularly used for diagnosis,  
 CC treatment or prevention of infection by Helicobacter pylori. Particularly  
 CC (Ia) and (IIa) are used in DNA, subunit or live vaccines. The method  
 CC identifies essential genes, including those that have homologs in many  
 CC species, so identified (A) should have a broad spectrum of activity. Many  
 CC gene-deficient cells can be screened quickly, in an automated process,  
 CC and the identified genes can be used for screening without purification.  
 XX  
 SO Sequence 916 AA:  
 4.4%; Score 206.5; DB 22; Length 916;  
 Best Local Similarity 19.3%; Pred. No. 1.6e-05;  
 Matches 202; Conservative 143; Mismatches 318; Indels 383; Gaps 51;

QY 459 T-LEPVIETVELTDLMD---ISPIEFSASMLI-----QDKLNLVAKARLHYDMP 506  
 DB 379 NGLVAVIYRIEVDKMYINDVIISGNQRTSDRIIRRELLGPKDKYNLTKLR----- 430  
 QY 507 DDRVLAINHDDGVNRSILIGRISDAVASAVARAIPDESENEVDLPERALANRKRPADY 566  
 DB 431 -----NSENSLR--LGFFSKVYIEKRY-----NSSLMDL----- 459  
 QY 567 QSKVPLVVEASDKPRDGI--GLMGSDTGTRLVTFEHLNIRDGYOAGAEIRLSED 624  
 DB 460 -----LVSEBGRGLOFGIGYSGGLM-----NSYSERNLFGTQGSMLAN 506  
 QY 625 --KGVKRIYATRP-----LSHPLMDQLRATLGYOQEVFGHSTNGFDLSTRYL- 669  
 DB 507 IATGGGRSPGMPKAGRMFAGNLSTNP-----RLF-----DSWYSTINLY 549  
 QY 670 -EHEISRIIIONG-----WNRTY-SLURYRLDKLTKQAPPETWQDLRYDFVNG-- 715  
 DB 550 ADYRISTYQYIOGGGFGVNVGRMLGNRTYHSLGYNLVTKLLGFSPPLYNRYSSVNEV 609  
 QY 716 KPSQFALLAGVAHKYVA-----DNLVNP-----MRGYROR-----YSLEV 751  
 DB 610 SPRGCTSPASVYIINRLSGKTPLPQESCSSPAITTSPEIRGIMDROCHTPTSSFTLDV 669  
 QY 752 G-----SSGLVSDANNAIRAGISGYVSFGD-----NAVSSNRAHOMTG----- 791  
 DB 670 SYDNTDTPYPRNGVIRS-----SYATMSGILPSSGTLNMGILGNVANKYKFAAYH 724  
 QY 792 -----IAGYIWSDNF--NHVPYRLRFAGGQOSIRGAHDSLSISDKGY 835  
 DB 725 HLGKYLIDLIARKKTGGGILFRYNTDDYLPNSTFYMGVTVYGRNGSVTPKDERGL 784  
 QY 836 LTGGQVLAVGTAEXYEFMK--DLRLAVFGDIG-----N 867  
 DB 785 WLGGDGIPTASTELSYGVLAAKMRILAMFPDFGLTETKTPRGSFFYNAPVTANPFDYG 844  
 QY 868 AYDKGFTNDI-KIGAGYGVNRASPVG 892  
 DB 845 VIGAGFERATWRASTGLOIEMISPMG 870

RESULT 27  
 ABG25558  
 ID ABG25558 standard; Protein; 371 AA.  
 AC ABG25558;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #25549.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS89745.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in



```

Db 315 VKEGQYRISDL-----IEIDNPVPLK--TLEKALKVKRKDVFNIEHLRADQ 362
OY 413 DL---IATRYENMVTLEVPEREQIONDVSEQSSSRTEPAQVDEST-LEPIETVE 468
Db 363 ILKTEIADKGYAFA---VVKPDLK-----DEKNGLVKVIYRIE 398
OY 469 LTGDLMDISPIEFASNLIDOKLNLVAAKARHLYDMPDR---VLAINHDDGVNRSILG 525
Db 399 VGD---MVHINDVIISGNQRTSDRI-----IRRELLGPKDKYNLKLKNSLSLR--LG 449
OY 526 RISDAVSAVARAILPDESENEVIDLPERTALANKTPADVQSKVPLVYFVASDKPRDG 585
Db 450 FFSKVKIEKRV-----NSLMDL-----LVSEEGRTG 478
OY 586 QI--GLGWSDTGTRLVYKFEHNLINRDGYOAGAELRLED--KKGYLYATKP----- 635
Db 479 QLOFGLGYSYGGLM-----NGSVSEKRLCTGQSMSLYANIATGGGKSYGMPKAGARM 534
OY 636 -----LSHPLNDQLRATLGYOEVEFGHSTNGFDLSTRTL--EHEISRSIIQNG-- 682
Db 535 FAGNLSLTPN-----RIF-----DSWYSTINLYADYRISYQIQGGGFGCVN 577
OY 683 -----NNRTY--SLRYRLDKLKTQAPETMODLPVDVFN--GKPSQELLAGVAVHK---- 730
Db 578 VGRMLGNRTHVSLGYNLNVTKLLGFSSPLYNRYSSVNEVASPQOSTPASIILNRLSG 637
OY 731 ---TVADNLVNP-----MRGYRQR-----YSLEVG-----SSGLVSDA 760
Db 638 RPLPVESSCSFGAITSPTPEIKCIMRDVHTPTTSSEFTLDVSTDNDDYFPRGVLFS- 696
OY 761 NMAIARAGISGYVSPGD-----NAVGSNRHQMOTG-----IOAG 795
Db 697 -----SATATSGLPSSSTGLMSMGLGNVNRKYYGKFAAYHHLQKLLIDLIAFKTQCG 752
OY 796 YVMSDNF--NHVYRRLFEFFAGDOSIRGYAHDSLSPISDKGYLGGVLAAGTAENYEFM 854
Db 753 YIFRNTDIDYLPINSTFYMGVYTVANGFNGSITTPKDEGLMLGGGIFTASTELSYGL 812
OY 855 K--DLRLAVFGDIG-----NAYDKGFTNDT--KIGAGVGY 885
Db 813 KAAMKRLAMFEDFGFLTEKTPTRGSEFYNAPTTANFKYGVVAGAFERATMRASGLGI 872
OY 886 RMASPVG 892
Db 873 EMISPMG 879

RESULT 29
AAV17183
ID AAV17183 standard; Protein: 925 AA.
XX
AC AAV17183:
XX
DT 03-AUG-1999 (first entry)
XX
DE H. pylori outer membrane polypeptide.
XX
KW Outer membrane polypeptide; OMP; vaccine; H. pylori infection; humoral;
XX cellular immune response.
XX
OS Helicobacter pylori.
XX
PN MO9921959-A2.
XX
PD 06-MAY-1999.
XX
PF 28-OCT-1998; 98WO-US22883.
XX
PR 17-DEC-1997; 97US-0993001.
XX
PR 28-OCT-1997; 97US-0959131.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
```

```

XX Alm RA, Ellis RW, Guild BC, Noonan BM, Smith D;
PI WPI: 1999-326698/27.
DR N-PSDB; AAX75802.
XX
PT Cellular vaccine against Helicobacter pylori
PS Claim 7; Page 243-247; 352pp; English.
XX
CC The invention relates to a vaccine for preventing or treating infections
CC by Helicobacter pylori. The vaccine contains at least one isolated
CC H. pylori polypeptide, or its fragments, in a carrier, where the
CC carrier is a Salmomella, Vibrio cholerae or Shigella vector containing a
CC nucleic acid encoding the H. pylori polypeptide. The vaccines induce a
CC humoral and cellular immune responses. The vaccines are used to treat or
CC prevent infections by H. pylori. Sequences AAX75779 to AAX75837 represent
CC (OMPs) AAV17160 to AAV17218.
XX
SQ Sequence 925 AA;
Query Match 4.1%; Score 194.5; DB 20; Length 925;
Best Local Similarity 19.9%; Pred. No. 0.00011;
Matches 192; Conservative 131; Mismatches 311; Indels 333; Gaps 49;
OY 116 TPLSLEELFMAQEST-----EMGINPNDYTP-EYQ-----GEOPSEVYV 153
Db 56 TPKEAQKNEAMQNETSQNQPKEMKVKYSISYVGLSYMSDMLANETAKRVGDWVDSKID 115
OY 154 PPTLEPKPKGLIKRLYARLPNDGVNKKVPRLKAQYSSQSG--ETSAGSSHQTEPYAN 211
Db 116 TAYALFENQGFQDVA--TEENG-----LEFHDEKARIAGVEIKGYTEKEKGKLSQ 169
OY 212 IKALEDITOESAMDLSIPRLQOTAL--VAARAVGYDDIDLSIRNSIGEVNVIINDLG 270
Db 170 MGAKGGTPEQKLE-----HAKTALKTALKEQGY-----GSV----- 203
OY 271 EPIYIDYRAVEVRGEGADKAFTTVAD--EVLPLIGDVPHGKYEYKKNLIEVASAEH- 326
Db 204 -----VEVTERKYSSEGALLIVFDVNRGDSIYIKQSIYESGDKRAKRVIESLSANKQ 254
OY 327 -----GYFQGR-----WDRSVD--VILPD--NTADVSLIYD 354
Db 255 RDEMGMMGLNDGKRLDLDLEYSRLIODVYMRGGLDAHISSPFLKTPDSTHAKLHYK 314
OY 355 T--GTQYRFEDEVVFETIDKTNQITDPPDKLPVKRELLBQLLVNMGGAVALNALSN 412
Db 315 VKEGIQYRISDL-----IEIDNPVPLK--TLEKALKVKRKDVFNIEHLRADQ 362
OY 413 DL---IATRYENMVTLEVPEREQIONDVSEQSSSRTEPAQVDEST-LEPIETVE 468
Db 363 ILKTEIADKGYAFA---VVKPDLK-----DEKNGLVKVIYRIE 398
OY 469 LTGDLMDISPIEFASNLIDOKLNLVAAKARHLYDMPDR---VLAINHDDGVNRSILG 525
Db 399 VGD---MVHINDVIISGNQRTSDRI-----IRRELLGPKDKYNLKLKNSLSLR--LG 449
OY 526 RISDAVSAVARAILPDESENEVIDLPERTALANKTPADVQSKVPLVYFVASDKPRDG 585
Db 450 FFSKVKIEKRV-----NSLMDL-----LVSEEGRTG 478
OY 586 QI--GLGWSDTGTRLVYKFEHNLINRDGYOAGAELRLED--KKGYLYATKP----- 635
Db 479 QLOFGLGYSYGGLM-----NGSVSEKRLCTGQSMSLYANIATGGGKSYGMPKAGARM 534
OY 636 -----LSHPLNDQLRATLGYOEVEFGHSTNGFDLSTRTL--EHEISRSIIQNG-- 682
Db 535 FAGNLSLTPN-----RIF-----DSWYSTINLYADYRISYQIQGGGFGCVN 577
OY 683 -----NNRTY--SLRYRLDKLKTQAPETMODLPVDVFN--GKPSQELLAGVAVHK---- 730
Db 578 VGRMLGNRTHVSLGYNLNVTKLLGFSSPLYNRYSSVNEVASPQOSTPASIILNRLSG 637
```





Db 709 PKDFGLWLGGOIFWASTELSYGLKAAKMRAMWFFDFGLFKPTPTGSEFFYNAPVTT 768  
 QY 867 -----NAVDFGFTNDT-KIGAGVGRWASPVG 892  
 Db 769 ANFKDYGIVGAGFERATWRASGTGLQIEWISPMG 801

## RESULT 31

AAR53758  
 ID AAR53758 standard; Protein; 793 AA.

AC AAR53758;  
 XX  
 DT 08-DEC-1994 (first entry)

DE H. Influenzae PAK 12085 D15 sequence.

DE H. Influenzae PAK 12085 D15 sequence.

KM Vaccine; passive immunization; vector; antiserum; diagnosis; D15;  
 KM OMP; outer membrane protein; NTH1.

XX Haemophilus influenzae PAK 12085.

OS  
 XX WO9412641-A.

PM 09-JUN-1994.

PD 23-NOV-1993; 93WO-CA00501.

PF 23-NOV-1992; 92GB-0024584.

PR (CONN-) CONNAUGHT LAB LTD.

PA Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;  
 XX Yang Y;

PI WPI: 1994-200269/24.

DR N-PSDB; AA066202.

XX Nucleic acid encoding D15 outer membrane protein - esp. of  
 PT Haemophilus influenzae, and related proteins, vectors, antisera  
 PT etc. useful in vaccines, for diagnosis and for passive  
 PT immunisation.

PS Disclosure; Fig. 1E; 161pp; English.

XX Outer membrane protein (OMP) D15 genes were isolated by screening  
 CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,  
 CC Eagan and Minn A, and the non-typeable (NTH1) strains SB33 and PAK  
 CC 12085. Nucleotide sequences were determined for the D15 genes  
 CC (AA066198-202) and the corresponding aa sequences were derived  
 CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,  
 CC free of other antigens and lipooligosaccharides, by recombinant DNA  
 CC methods using the isolated genes.

XX Sequence 793 AA;

XX Query Match 3 9%; Score 184.5; DB 15; Length 793;  
 XX Best Local Similarity 20.0%; Pred. No. 0.00043;  
 XX Matches 167; Conservative 129; Mismatches 326; Indels 211; Gaps 41;

QY 227 LNSGIPRLRQFALVAAVAVGYDIDLSIRNSI---GEV-DVLIHDLGPVYIDYRA--- 279  
 Db 32 VOGDLEQIIRASLPYRAGHGVTDNDVANIYRSLFVSGRFDVKAHQEGVLYVSVYAKSI 91

QY 280 ---VEVRGEGA-DDKAFVTVADEVPLLIGDVFHNGKYEETKKULIENASAEHGFQGRWLD 335  
 Db 92 ISDVKIKGNSVLPTEALQNDLANGEFKVDVL---IREKLNEFAKSYKEHVASGRY-N 146

QY 336 RSVQYI---LPDNTADVSLIVDTGYQYRDEVEVFTIPDKTQULTTDPKLVKRELEQ 392  
 Db 147 ATVEITVNTLPNNRAELIQINEDDKAKLASLTF---KGNE-SVSSSTLQEOHLEQDP 200

QY 393 LLTVNMGFAV-----NLQAVRALSNDLIATRYFNVMVTEIVPEREQIQNDQVSFEQS 445

Db 201 SMMKLMGNKKEFGAOFEDLOAIR-----DYLLNNGYAKAQTTKTDVOLNDEKTYVNV 252  
 QY 446 SSSRTEPAQVD-----ESTLEPIYETVELLDGI-LMDISPIETESANLIQDK 491  
 Db 253 TIDVNEGLQYDLRSARITIGNLGMSAELEPLLSALHLNDPFRSDIDVE---NAIKAK 308

QY 492 LNLVAAKARHLIDMPD---DRVLAINHDDGVNSIL-----GRISDAVSAAVAILP 540  
 Db 309 LGERYGNTTVNSVDPEDDANKTLAIFVYDAGRLLTVROLRFEGNTVVSADSTLRQERQ 368

QY 541 DES---ENEVIDLP---EPT---ALANKRTPAD-----YQSKKPLVLYFVASDK 581  
 Db 369 QEGTWINSQVELKTLIDRTGFPEYENKIDPLNGSNDDEVYVYK-----E 417

QY 582 PRDQOI---GLWGSDDTGRVLYKE-EHNLJ-----NRDGYOAGELRLSE---DK 625  
 Db 418 RNTGSINFGIGYGESGISYQTSIKQDNFLGTGAAVSIAGTKNDYGSVNLGYTEPYFTK 477

QY 626 KGYKLYATKPLSHPLNDQLRATLGYQOEVRGHS-TNGFDLSTR-----LEH---EISRS 676  
 Db 478 DGVSILGNIFFEYNDNSKSDTSSNYKRTTYSNVTLLGEPVNNSSYYVGLGHTYKISNF 537

QY 677 IIONGGMNRT---YSLRYRLDKLTKQAPPEWODLPDVFVNGKPSQALLAGVAVHRTVA 733  
 Db 538 ALE---YNRNLTYIQSMFKNGIKTN-----DFDFSGW----- 568

QY 734 DNLVNPGRGYRORYSLEVGSSGLV---SDANMAIARAGISGVYSGFDNAYGSRHAHQ-- 787  
 Db 569 -NYNSLNRGYFPTKGVKASLGGRVTTIPGSDNKYKLSADVQGFYPL-----DRDHRWV 620

QY 788 MTGGIGAGYIWSDFNHNVPYRLREFAGGDSIRGYANDLSL---ISDKGLT----- 837  
 Db 621 VSAKASAGYANGFGNKRRLPYQYTTAGTIGSLRGFAVGSIGPNAIYAEHNGTFNKKISSD 680

QY 838 --GGQVLAVGTAE--YNYEFMKD-----LRLAVFGDIGNAY----- 869  
 Db 681 VIGGNMATTASAEIIVTFPVSQDSQNTVTSLEFVDAASVWNTKMKSDKNGLESYVLKDL 740

QY 870 -DKGFTNDTKIGAGVGRWASPVQGVAVDVATGVK--EEGNPIKLHFFIGTPF 919  
 Db 741 PDYGKSSRIRASTGVGVQWQSPIGLPLFVSYAKPIKYYENDVDQFQFSGISGF 793

## RESULT 32

AAR53757  
 ID AAR53757 standard; Protein; 792 AA.

AC AAR53757;  
 XX  
 DT 08-DEC-1994 (first entry)

DE H. Influenzae SB33 D15 sequence.

DE H. Influenzae SB33 D15 sequence.

KM Vaccine; passive immunization; vector; antiserum; diagnosis; D15;  
 KM OMP; outer membrane protein; NTH1.

XX Haemophilus influenzae SB33.

OS  
 XX WO9412641-A.

PM 09-JUN-1994.

PD 23-NOV-1993; 93WO-CA00501.

PF 23-NOV-1992; 92GB-0024584.

PR (CONN-) CONNAUGHT LAB LTD.

PA Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;  
 XX Yang Y;

PI WPI: 1994-200269/24.

DR

```
DR N-PSDB; AA066201.
XX
XX Nucleic acid encoding D15 outer membrane protein - esp. of
PT Haemophilus influenzae, and related proteins, vectors, antisera
PT etc. useful in vaccines, for diagnosis and for passive
PT immunisation.
XX
PS Disclosure; Fig. 1D; 161pp; English.
XX
CC Outer membrane protein (OMP) D15 genes were isolated by screening
CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
CC Eagan and Minn A, and the non-typable (NTHi) strains SB33 and PAK
CC 12085. Nucleotide sequences were determined for the D15 genes
CC (AA066198-202) and the corresponding aa sequences were derived
CC (AAR33754-58). D15 OMP can be produced easily and on a large scale,
CC free of other antigens and lipooligosaccharides, by recombinant DNA
CC methods using the isolated genes.
XX
XX Sequence 792 AA:
SO
Query Match 3.8%; Score 178; DB 15; Length 792;
Best Local Similarity 20.2%; Pred. No. 0.0012;
Matches 160; Conservative 118; Mismatches 306; Indels 210; Gaps 39;
QY 263 DVIHDLGEVYIDYRA-----VEVRGEG-ADDKAFITVADEVPLIGVFNHGYETK 315
DB 72 DYKAHQEGDVLVYVAVKSIISDVKIKGNSIIPPEALKQMLDNGFKVGDIL-----IREK 127
QY 316 KNIENASAHGEYDFGFWMLRSYDVI---LPDNTADVSLIYDICTQVRFDEVFETDPK 372
DB 128 LNFPAOSVKEHYASVGRY-NATVEPIVNTLPNNKAETLIQINEDDKAKLASLTF-----K 181
QY 373 TNOITDPPKLPVKRELLLEOLLTVNMGEAY-----NLQAVRALNDLIATRFENNVNT 425
DB 182 GNE-SVSSSTLQOMELQPSMWKMGKNGKEGAQFEKDQAIR-----DYLNNGYA 232
QY 426 EIVFPEREQIQNDQVSESSSSRTPEQVD-----ESTLEVIETVELTDG 472
DB 233 KAQITADVOLNDEKTKVNTITDINEGLQYDLSARIITIGLGSASLEPLLSALHNDT 292
QY 473 I-LMDISPIEFASNTLQDKLNVAAKARHLXMDP---DRVLAITHHDGVNRSIL--- 524
DB 293 FRRSIDIADE---NAIKAKLGERGYGNTVNSVDPEDANKLATITFVVDAGRRLLVHQ 348
QY 525 ---GRISDAVSAVARAILPDES---ENEVIDP---ERT---ALANKRTPAD----- 564
DB 349 LREGNTVASDSTLRQEMRQEGTWTWNSQVLELGRIDLTGTFFVEYENKIDPLNGSND 408
QY 565 ---VYOSKVPPLVYVASDKPRDQI--GLWGSDTGTRLVTKF-EHNLV----- 608
DB 409 VDVVVYVK-----ERNTGSINFGIGYGTESGISYQASVKQDNELGTGANVSIAG 457
QY 609 NRDDGYOAGELRLSE---DKKGVKLYATKPLSLPDLQRLATIGYQOEYFGHS-TMGFDL 664
DB 458 TKNDDGTSVNLGTEPYETFKDVSIGGNVFEFENYDNKSDTSSNRYKRTYGSVWTIGFPV 517
QY 665 SFRY-----LEH---EISRSIIQNGGMNRT---YSLRYLRLDKLQAPRPTWDDLPDEV 713
DB 518 NENNSYVYGLGHTYKNISFPALE---YNNRLYIQSMKFKGNGKTKN-----DFDRS 565
QY 714 NCKPQOEALLAGVAHVKTVAADLVNPMRGYRORYSLEVGSSGLV---SDANNAIARAGT 769
DB 566 FGM-----NYSNLNRGYPFPTKGVKASIGGRVTLPGSDNRYKYLISADV 607
QY 770 SCVYSPGONAYGNSRAH---QMTGGIOAGYIMSNFNHNVYRLRFEGGDSITGYAHDSTL 827
DB 608 QCFYPL-----DKDHLVWVSAKASAGANGFNKRRLPEFYQTYTAGIGLSLNGFAYGSI 660
QY 828 SP-----ISDKGYLTGCGOVLAVGTAE--YNEFPKMD-----LRLAVPEDIGN 867
DB 661 GPNALTYQGGNNKFNKISSD-VIGGNAITATASABELIVPFPYSDKSQNTVRFSLFVDAAS 718
QY 868 AY-----DKGFTYNDTKIGAGYGVWMASPVQGVRRVAVATGVK--EE 905
```

```
DB 719 VWNTRKWSKDKNGLESNVLKDLPDYGSSSRTBRASGVGFQWQSPGQVYFSAKPIKTYEN 778
QY 906 GNPILKIHFFICTPFF 919
DB 779 DDVEQFOFSTIGSF 792
RESULT 33
AAV44390
ID AAV44390 standard; Protein; 813 AA.
XX
AC AAV44390;
XX
DT 14-MAR-2000 (first entry)
XX
DE M. catarrhalis BASB027 polypeptide #1.
XX
XX BASB027; OMP85; outer membrane protein; otitis media; treatment;
KM diagnosis; bacterial infection.
XX
OS Moraxella catarrhalis.
XX
XX W09963093-A2.
XX
XX 09-DEC-1999.
XX
XX 31-MAY-1999; 99WO-EP03822.
XX
XX 03-JUN-1998; 98GB-0011945.
XX
XX 08-MAR-1999; 99GB-0005304.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Vinals-Bassols C;
XX
XX WPI; 2000-105700/09.
XX
XX N-PSDB; AA229550.
XX
PT Novel BASB027 polynucleotide and polypeptides from Moraxella
PT catarrhalis useful for treating M. catarrhalis infection such as otitis
PT media
XX
PS Claim 1; Fig 3; 109pp; English.
XX
XX The present sequence is BASB027 polypeptide, which shows significant
XX homology to Neisseria meningitidis OMP85 outer membrane protein. It is
XX encoded by DNA obtained from chromosomal DNA library of Moraxella
XX catarrhalis strain Mc2931 (ATCC 43617). BASB027 polynucleotide and
XX polypeptide can be used for diagnosis and staging of disease, determining
XX susceptibility to a disease and to prepare medicaments for treating M.
XX catarrhalis infections, especially otitis media. The BASB027 DNA can be
XX used as probe for screening of genetic mutations, serotype, taxonomic
XX classification or identification. BASB027 agonists, antagonists and
XX antibodies may be used to prevent and/or treat bacterial infections.
XX
SO Sequence 813 AA:
Query Match 3.7%; Score 176.5; DB 21; Length 813;
Best Local Similarity 18.7%; Pred. No. 0.0016;
Matches 164; Conservative 119; Mismatches 338; Indels 257; Gaps 36;
QY 165 IKRLYARLFMDGVNKKPRKAKFYQSSQGETSAIGSSHOKTEPYANIKALEDTQESA 224
DB 70 VKALYATGNFSDVQVYHQEGRITTYVTERPLAINEFGRNRLIPKELQGLKNAGIAYG 129
QY 225 MDLNGSIPRLROTALVAAR-AVGYYDIDLSITNSIG---EYDVIHDLGEVYIDYRA 279
DB 130 QPLKQATVOMIETELNQTYSQGYNTETIVKOTMLDGNRVKIDMTFAE-GKPA-----RV 184
QY 280 VEYRGAGADKAFITVADEV-----PLTIGDVFHNGYETKKNLIENASAEH---GY 328
DB 185 VDNINIGNHFSADLIDVLAIKDNKINPLSKADRYTQELVYS---LENLRAKYLINAGF 241
```

```
QY 329 FDGHWLDRSVQVILPDNTADVSLIYDTGYRFEDEVVEFTIDPKTNQLTDPDKLPVRE 388
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 242 VREIIRDAKININEDKNRIFVEISLHGEQYRFGQTOFL-----GNLTQTA 288
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 389 LLEQGLTVNMGEAVNLQAVRALSNDLIATR-----YENMVN--TELVFEREQIQNDQV 440
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 289 ELLEALKFKAEEGFS-QAMLEQTTNNISTKFGDDGYYYAQIRPVTRI-----NDE- 337
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 441 SFEQSSSRTEPAQVDESTLEPIVETVELDGIIMDISPIEFASNLIDKLVAKAR 500
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 338 -----SRIVDEYIYIDPHVPIY-----VRRIINFTGNFKYQDE--VLRRMR 376
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 501 HLYDMPDDRVLAINHDDGVNRSILGRISDAVSAVARAILPDESENEYIDLPERITALNRK 560
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 377 QL-----EGALASNOKIQLSR-----ARLMRTGFKKHYIVD--TRPVNSP 415
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 561 TRPDVYQSKKVPILYVAVASKPR-DGQIGLGWGSQDGTGLVTKFEHNLINRDGYQAGAEI 619
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 416 DOVDVN-----FVVEQPSGSSSTIAAGYSQSGGVTFQFDVSQNNFMGTGKHVNASF 466
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 620 RLSEDKKGVLYATKPLSHPLNDQLRATLGYQOEVEFGHSTNGFDLSTRLEHISRSIIQ 679
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 467 SRSETREYVSLGKMTNRY-----FTVNGVSQSLSGYRKTKYDNKNISNYVD 513
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 680 NGGMNRTYSLRYRLD-----KL-----KTQAPPETWQ 706
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 514 SYGGSLSYG--YPIDENQIRISFGLNADNRTKLHGGRMGISNVKQMLADGKIQVDN--N 568
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 707 DLDPVDVNGKPSQEALLAGVAVHKTVADNLVPMRGYRQRYSLIEVGSSGLVSDANMAIAR 766
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 569 GIP-DKHKHYTYTMAILGW--NYSSLDPRVPFTQGMSSHVDLTVG----- 610
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 767 AGISGVYSFGDNAYGNSNRHQ-----MTGSIQAGYIMSDNFMNVPRLRFEAG 814
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 611 -----FGD-----KTHQKVYVQGNIRYPRITKKSVLKGVAKLGGVNLPEYENFYAG 656
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 815 GQOSIRGVYHADSLSPISDKGYLT-----GGQVLAVGTAE-----YNEEFMK 855
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 657 GYGSVAGYDQSSSLGPRS-QAVLTARQGQTTLGEVVGNAALAFGSELLPLPFKGDWID 715
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 856 DURLAVFGDIGNAYD-KGFTTNDT-----KITAGY 883
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 716 QVRPIVIEFGGVFTDTGMDKQITDITQFQDPQATAEQNAKANRBLIDQDKQLRYSAGV 775
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 884 GVRMASPVQGVADVATGV--KEESNPILHFFIGP 919
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 776 GATWYTPIGPLSISTAKPLNKKONQDTTVOFOIGSVF 813
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 34
AAV44391
ID AAV44391 standard; Protein; 813 AA.
XX
AC AAV44391;
XX
XX 14-MAR-2000 (first entry)
XX
XX M. catarrhalis (ATCC 43617) BASB027 polypeptide #2.
XX
XX BASB027; OMP85; outer membrane protein; otitis media; treatment;
XX
XX diagnosis; bacterial infection.
XX
XX Moraxella catarrhalis.
XX
XX M09963093-A2.
XX
XX 09-DEC-1999.
XX
XX 31-MAY-1999; 99WO-EP03822.
XX
XX 03-JUN-1998; 98GB-0011945.
XX
XX 08-MAR-1999; 99GB-0005304.
XX
```

```
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Vinals-Bassols C;
XX
XX WPI: 2000-105700/09.
XX
XX N-PSDB; AA229551.
XX
PT Novel BASB027 polynucleotide and polypeptides from Moraxella
PT catarrhalis useful for treating M. catarrhalis infection such as otitis
PT media
XX
XX Claim 1; Page 102-104; 109pp; English.
XX
XX The present sequence is BASB027 polypeptide, which shows significant
XX homology to Neisseria meningitidis OMP85 outer membrane protein. It is
XX encoded by DNA obtained from chromosomal DNA library of Moraxella
XX catarrhalis strain MC2931 (ATCC 43617). BASB027 polynucleotide and
XX polypeptide can be used for diagnosis and staging of disease, determining
XX susceptibility to a disease and to prepare medicaments for treating M.
XX catarrhalis infections, especially otitis media. The BASB027 DNA can be
XX used as probe for screening of genetic mutations, serotype, taxonomic
XX classification or identification. BASB027 agonists, antagonists and
XX antibodies may be used to prevent and/or treat bacterial infections.
XX
SQ Sequence 813 AA;
XX
Query Match 3.7%; Score 176.5; DB 21; Length 813;
Best Local Similarity 18.7%; Pred. No. 0.0016;
Matches 164; Conservative 119; Mismatches 338; Indels 257; Gaps 36;
QY 165 IKRLYALFMDGVNKKVRLAKFYQSSQSGTSAIGSSHQKTEPYANIKALEDITQESA 224
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 70 VKALYATGNESDVQVYHQEGRIIYQYTERPLIEMINEGNRLIPKESLOELKNAGLAVG 129
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 225 MDLNGSIPRLQOTALVAVR-ANGYDIDLSTIRNSIG---EVDYIIHIDGEPYIDYRA 279
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 130 QPLKQATVQMIETELTMQYISQGYNTETVTKQMLDGNRKVDIMTPAE--GKPA---RV 184
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 280 VEVREGACDAKFAFTYADEV-----PLLIGDFVHNGKYEYTKKNLIENASAEH--GY 328
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 185 VDINITIGNQHFSDADLIDVLAIDKNKINPLSKADRYQELKLYVS---LEMLRAKIYNAGF 241
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 329 FDRWMLDRSDVILPDNTADVSLIYDTGYRFEDEVVEFTIDPKTNQLTDPDKLPVRE 388
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 242 VREIIRDAKININEDKNRIFVEISLHGEQYRFGQTOFL-----GNLTQTA 288
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 389 LLEQGLTVNMGEAVNLQAVRALSNDLIATR-----YENMVN--TELVFEREQIQNDQV 440
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 289 ELLEALKFKAEEGFS-QAMLEQTTNNISTKFGDDGYYYAQIRPVTRI-----NDE- 337
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 441 SFEQSSSRTEPAQVDESTLEPIVETVELDGIIMDISPIEFASNLIDKLVAKAR 500
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 338 -----SRIVDEYIYIDPHVPIY-----VRRIINFTGNFKYQDE--VLRRMR 376
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 501 HLYDMPDDRVLAINHDDGVNRSILGRISDAVSAVARAILPDESENEYIDLPERITALNRK 560
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 377 QL-----EGALASNOKIQLSR-----ARLMRTGFKKHYIVD--TRPVNSP 415
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 561 TRPDVYQSKKVPILYVAVASKPR-DGQIGLGWGSQDGTGLVTKFEHNLINRDGYQAGAEI 619
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 416 DOVDVN-----FVVEQPSGSSSTIAAGYSQSGGVTFQFDVSQNNFMGTGKHVNASF 466
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 620 RLSEDKKGVLYATKPLSHPLNDQLRATLGYQOEVEFGHSTNGFDLSTRLEHISRSIIQ 679
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 467 SRSETREYVSLGKMTNRY-----FTVNGVSQSLSGYRKTKYDNKNISNYVD 513
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 680 NGGMNRTYSLRYRLD-----KL-----KTQAPPETWQ 706
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 514 SYGGSLSYG--YPIDENQIRISFGLNADNRTKLHGGRMGISNVKQMLADGKIQVDN--N 568
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 707 DLDPVDVNGKPSQEALLAGVAVHKTVADNLVPMRGYRQRYSLIEVGSSGLVSDANMAIAR 766
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

Db 569 GIP-DEFHDTTNTALLGW---NTSSLDRPVFPFGSMHSVDLTVG----- 610  
QY 767 AGISGVYSGFDNAYGSRNRAHD-----MTGGIQAQGIWSDNENHVPYRLRFPAG 814  
Db 611 -----FGD-----KTHQKVVYQGNITVRPFIKKSVLRTGRTALGIGNNLPFIENETIAG 656  
QY 815 GDOSIRKTYANDLSPIIDKGYLT-----GGQVLAVGTAE-----YNYEFMK 855  
Db 657 GYGSVNRGYDSSLSLGRS-QAVLTFARQOQTTLGEVGNALATFGSELLPLPKGDWID 715  
QY 856 DLRLAVGDIQGNMVD-KGFINDT-----KIGAGV 883  
Db 716 QVRPVITIEGGVFDTTGMDKQITDLTQFKDPQATAEOMAKAANRPLLTDOKLRYSGAV 775  
QY 884 GVRMASPVGOVRVDVATGV---KEEGNPDKLHFFIGTPE 919  
Db 776 GATWYTPIGPLSISYAKPLKKKQNDQTDYQFOIGSVF 813  
RESULT 35  
ABBS5843  
ID ABBS5843 standard; Protein; 2748 AA.  
XX  
AC ABBS5843:  
XX  
DT 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 3321.  
DE Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN MO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI: 2001-656860/75.  
XX N-PSDB: ABL02946.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclosure: SEQ ID NO 3321; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
CC sequences (ABBS57737-ABBS72072).  
CC (ABBS57737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
XX  
SQ Sequence 2748 AA;

QY 9 NRSEFPVALAAYLPLMTSOALAOQNNPANIINHVPADTAI-----NQAKAGNPVLLTPE 64  
Db 1217 SRAIEPLPLAP-----PAEVLVYPLGSAELLSSEPELVVSPSIALPE 1261  
QY 65 QIQARLNAAGIN---AKPOSALDVVNFDDOSPISIRIGOSPPLGLDMVIEETPLSL 120  
Db 1262 KPEPMENATEIEIDEMVSPTVEIVPTTEDLHQAEVVEVETARP---VSAVEEVKPI-- 1315  
QY 121 EELFAQESTENGINPNDIYPEYQGE-QPNSEVVVPPITLPEPKPGILKLYARLNDGYNK 179  
Db 1316 -----EDSTAIDKSPDISPIEGEHSQSEKETPESEVE---VGLDSPNADVITTSVEV 1366  
QY 180 VPRLLAKAFYOSSGSETSAIGSHQKTEP-----YANIKAALEDITQ----- 221  
Db 1367 IP-----TSTAAYITSDGASTEMPEPKYICETAPPAEIEIQNETVETLSYADTT 1418  
QY 222 --ESAMDLNGSIPRLROTALVAAVAGYYDLSIIRN---SIGEVDTI----- 266  
Db 1419 LPEEIVSQDGNLEASQTA-----SIDVTFQANLNMEKSGCEVAVDRESKED 1468  
QY 267 HDLGEPIYIDYRA-VEV-----REBGADKAFITVADEV 300  
Db 1469 ETLMEPIAVDVAVVEVYVPOCAGVELGYSEPEEATPATNLTSGENPPEEV-SELVDYEP 1527  
QY 301 LLIGDVFHGGKYEETKKNLIENASAEHGFQGRMLDRSY-----DYILPDNADY 349  
Db 1528 RNPEEIHITDSFEKDSOKSSDAVALHE--DWTKEEKKVDPPQVAEVAAPDIITLDPSSLV 1565  
QY 350 SLIYDT-----GTQYRPD-----EVVFTIDRKTNLTQTPD----- 381  
Db 1586 ELSPDTQAMIGNIAPPDILEPESIVIEQVPVVKPBTIYESDPEAIDHTALLEPSVS 1645  
QY 382 -----KLVPRELLEQLITVNMGEAYVLQAVRALS-NDLIATRYFNMYN----- 424  
Db 1646 AEYODKVESAOQVSEKNSVEEDSLKYSIGERGEIISMDPTPADAVSEQNFRVEDPTTS 1705  
QY 425 -----TEIVPEREQIQNDQVSFEQSSSRTEPAQVDESTEPIYE--TVBLTGILMDI 477  
Db 1706 TESTAVEVIASESINTVVEPVSAQSPF-----IDSTPEEVTQCSAVNSTPEVASA 1758  
QY 478 SPIEFSASNLIODKLNLY-----AKARHLVDPDDRVLAINHDDGYNRSIGRIDAV 531  
Db 1759 APIQ-----VDQKIQPVHVELVQSIEASDQESPKD--TAQVSDSVDAIILQVSETD 1810  
QY 532 SAVARAILP---DESENE-----VIDLPERTALNKRTPADVYOS--KKVP----- 572  
Db 1811 SILVEPSSPVKQSEDSQEPGSAQNVLDV-----GPVEMDQSAPEVEPETDPA 1860  
QY 573 -----LVYFVASDKPRQOQIGLWGSDDGTGLVYKFEHNLINRQCYQAGALRL 621  
Db 1861 VEPSVPVEEVEVIEIQAPEKISDSVERTLDSLQNTDLVAEAS--VDTPAENMGPIGT 1918  
QY 622 SEDKKGVKLYATKPLSHPLNDQLRATG-----YQGEVFGSTGDFLSTRTEHEHSRS 676  
Db 1919 SELEKELLYVT-----TIDQPSAVLEPLVLICQAE-----STN-----LDHELPE 1961  
QY 677 TIONGNNRTYSLRYRDLKLTQAPE--TWODLPEVDFVNGKPSGEALLAGVAAHKYVADN 735  
Db 1962 LVPOG-----QSVPASLPTNLTIQTQPTD--LETQPSMELITQ---TETVDAE 2004  
QY 736 LVNPM 740  
Db 2005 MKRPV 2009  
RESULT 36  
AAR53755  
ID AAR53755 standard; Protein; 797 AA.  
XX  
XX  
AC AAR53755;  
XX  
DT 08-DEC-1994 (first entry)

```

XX DE H. Influenzae b Eagen D15 sequence.
XX KW Vaccine: passive immunization; vector; antiserum; diagnosis; D15;
XX KW OMP; outer membrane protein; H1b.
XX OS Haemophilus influenzae type b Eagen strain.
XX PN W09412641-A.
XX PD 09-JUN-1994.
XX PF 23-NOV-1993; 93WO-CA00501.
XX PR 23-NOV-1992; 92GB-0024584.
XX PA (CONN-) CONNAUGHT LAB LTD.
XX PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
XX PI Yang Y;
XX DR WPI: 1994-200269/24.
XX DR N-PSDB; AA066199.
XX CC Chromosomal libraries of H. influenzae type b (H1b) strains Ca,
XX CC Eagan and Minn A, and the non-typeable (NTH) strains SB33 and PAK
XX CC 12085. Nucleotide sequences were determined for the D15 genes
XX CC (AA066198-202) and the corresponding aa sequences were derived
XX CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
XX CC free of other antigens and lipooligosaccharides, by recombinant DNA
XX CC methods using the isolated genes.
XX SQ Sequence 797 AA;

Query Match 3.6%; Score 171.5; DB 15; Length 797;
Best Local Similarity 19.8%; Pred. No. 0.0034;
Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

QY 263 DVIHTDCEPYIYDRA-----VEVRGECA-DKKAFTTVADEVPLLGDFVHNGKETK 315
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 72 DVKAHQEDVLAVSVYAKSIISDVKIKGNSVPTFALKONLDANGCFKYGDVL----IREK 127
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 316 KNLEMASAEHGYFDGRMLDRSVDI---LPDNTADVSLIYDTGQYRDEVVFFETIDPK 372
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 128 LNEFAKSYKEIHASVGRY-NATVEPIVNTLPNNRAEILIQINEDDKAKLASLTF-----K 181
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 373 TQNLTTDPDKLPVKRNLLEQLTLTNMGEAY-----NLQAVRALSNDLIATRYFMVNT 425
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 182 GNE-SVSSSTIQEQMELQPDSSWMKLMGNKFEGAEQFKDLQSR-----DYILNCGYA 232
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 426 EIVPEREQIONDOVSFEQSSSRTEPAQV-----ESTLEPIVETVELTDG 472
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 233 KAOITTKTDVQINDEKTKVNTIDVNEGLOYDLRSARIICNLGMSAELEPLLSAHLNDT 292
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 473 I-LMDISPIEFSASNLIDDKLNLVAKARHLYDMPD---DRVLAINDDGVRNSIL--- 524
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 293 FRRSIDIADVE-----NATAKAGERGYSATVNSVPDPDANKTLITLTVDAGRRLTVRQ 348
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 525 ---GRISDAVASAVARAILPDES---ENEVIDLP-----ERT---ALANRKTTPAD----- 564
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 349 LRFGNTVSAOSTLRQEMRQEGTMVNSQLVELGKIRLDRGFEFTVENRDRIDPIGNSDE 408
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 565 ---VYQSKKVPILYFVNASDKPRDGOI--GLGWSGDTGRVLTQK- EHNLI----- 608
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 409 VDVVYKVK-----ERNTGSIINGIGYGTESGISYQASVYKODMFLGTGAIVSIAG 457
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

```

```

QY 609 NRDGIQAGAEILRISE---DKKGVKLYATKPLSHPLMDQLRATIGYQOEYFGHS-TNGFDL 664
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 458 TKNDYGTSVNLGTYEPTFKDGVSLGCVNFFENYDNKSDTSSNRYKRTYGGSNVTIGFVY 517
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 665 STRT-----LEH---EISRSIIONGGMNRT---YSIRYRLDKTKQAPETWQDLVDVY 713
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 518 NENNSTYVGLGHTYKISNPALE---YNRNLYIQSKFNGIKTN-----DFDPS 565
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 714 NGKPSQELLAGVAHVKTADVADNLVNPMPRGYRORYSLEVGSSGLV---SDANMAIARAI 769
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 566 FGV-----NYSNLNRCYFPPTKGVKASLGGVTTIPGSDNKYKLSADY 607
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 770 SGVYSRGDNAVGSNRH--QMTGGIAGYIWSNFMHVRRLRFFFGQDSIRGVAHDSL 827
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 608 QGFYPL-----DRDHLWVYSKASAGYANGFGNKRPLPFYQTYTGAGISLGFAYGSI 660
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 828 SP-----ISDKGYLTGGOVLAVGTAE--VYEFMKD-----LRLAVF 862
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 661 GPRALIYAEGNGSGTGFKKISSD--VIGNMAIATASAEILYPTFPVSDKSQNTVTSILF 718
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 863 GDIGNAY-----DKGFTNDTKIGAGVGYRMASPVGYQVRDVATGV 902
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 719 VDAASVWNTKMKKSDKNGLSDVLRPLPDYKSSRIRASTGVGFQWQSPIGPLVFSYAKPI 778
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 903 K--EEGNPTKLHFFIGTPTF 919
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 779 KRYENDVBEQFOFISIGSF 797

RESULT 37
AAR53756
ID AAR53756 standard; Protein; 797 AA.
XX AC AAR53756;
XX DT 08-DEC-1994 (first entry)
XX XX
XX DE H. Influenzae b Minn A D15 sequence.
XX KW Vaccine: passive immunization; vector; antiserum; diagnosis; D15;
XX KW OMP; outer membrane protein; NTH1.
XX OS Haemophilus influenzae type b Minn A strain.
XX PN W09412641-A.
XX PD 09-JUN-1994.
XX PF 23-NOV-1993; 93WO-CA00501.
XX PR 23-NOV-1992; 92GB-0024584.
XX PA (CONN-) CONNAUGHT LAB LTD.
XX PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
XX PI Yang Y;
XX DR WPI: 1994-200269/24.
XX DR N-PSDB; AA066200.
XX CC Chromosomal libraries of H. influenzae type b (H1b) strains Ca,
XX CC Eagan and Minn A, and the non-typeable (NTH) strains SB33 and PAK
XX CC 12085. Nucleotide sequences were determined for the D15 genes
XX CC (AA066198-202) and the corresponding aa sequences were derived

```

CC (AAV53754-58). D15 OMP can be produced easily and on a large scale.  
CC free of other antigens and lipooligosaccharides, by recombinant DNA  
CC methods using the isolated genes.

**SQ Sequence 797 AA;**

Query match 3.68; Score 169.5; DB 15; Length 797;

Best Local Similarity 19.8%; Pred. No. 0.0047;

Matches 158; Conservative 120; Mismatches 306; Indels 215; Gaps 39;

QY	263	DVILHDLGPVYIDIRA-----	-VEVBEGA-DKAFITVADEVPLIGVFFHHKYEYK	313	
Db	72	DVKAHQGGVLLVSVVAASIIISDKIKGNSVLPTEALKQNLDANGFKVGDL----	-IREK	127	
QY	316	KLILINASAEHGYFEGRMIDRSVDY---	-LPDNTADVSLIYDTGTOYRFDEVEFTIDPK	372	
Db	128	LMEFAKSVKEHVAASGRV-NATVEIVTITLNNRAELIIQINEDDKKLASLT-----	-K	181	
QY	373	TNQLTTDDPKLVPKRELLSQLTTVMGEAY-----	-NLAVALRSLNDLIATRYFNMYNT	425	
Db	182	GNE-SVSSSTLQEQMELQPDMSWMKLMGKPFGAQEKDQSGIR-----	-DYLLNNGYA	232	
QY	426	ELVFPEREDIQNDQVSEFQSSSRPEPQVD-----	-ESTLEPIETVELTDG	472	
Db	233	KQITKTQDLQDETKKYNVIDNEGQYDLRSARIIGNLGMSAELEPLLSALHNDT	292		
QY	473	I-LMDSPIEFASNLIDOKLNLVAKARHLIDMPD-----	-DRVLAINHDDGVNSIL-	524	
Db	293	FRRSDIADVE-----	-NATRAKIGERGYSATVNSVPDEDDVAKTTLATILVVDAGRLTVRQ	348	
QY	525	---GRISDAVASAVARAILPDES---ENEVIDLP-----	-ERT--ALANKRTPAD-----	564	
Db	349	LRFEGNTVSADSTLHQERKQEGTWYNSQLVELDKIRLDRIGFEFYEYENRIDPIGNSDE	408		
QY	565	---VYQSKKVPILYFVNASDKPRDGOI--GLGMSDGTGRLVTKF--EHNLI-----	608		
Db	409	VDVVYKVR-----	-ERNSTGSINFGIGYGTESGISYQASVYKQDNFELGTGAANSIAG	457	
QY	609	NEDQVQACAEELRLSE---DKGKVKLYARKPLSHPLNQRLATLGTQOQVGVGHS--TNGFDL	664		
Db	458	TKNQYGTSVNLGTYPTTKQGVSLGCVNFFENYDNSKSDTSSNYKRTYTGSVNTTIGEPV	517		
QY	665	STRT-----LEH---EISRSIIIONGWNRT--YSLRYRLDKLKTQAPETWODLPVDFV	713		
Db	518	NENNSYVYGGLGHTYKKSINFALE---YNRNLYIQSMFKFGKIGITN-----	-DPDFS	565	
QY	714	NGKPSQEQELLAGVAVHKTVADNLVPMKRGYQRYSLVEYGSSGLV---	-SDANNAIARAGI	769	
Db	566	FGW-----	-NYSNLNLGYPPTKGVKASLGGRYTIPGSNKKYYKLADY	607	
QY	770	SGVYSFGDNNAVGSNRAH--OMTGGIQQGYIMSDNFNVPYRLRPFAGDGSINGYAHDSL	827		
Db	608	QGFYPL-----	-DRDLHLMVNSAKASAGYANGFGCNKRKLPEYQYTTAGIGISLGLGFAVGS	660	
QY	828	SP-----	-ISDKGYLTGGQVLAVGTAE--YVNEFPKMD-----	-LRLAVE	862
Db	661	GPMNIIAYEYNGSGTGTFKKIISD--YIGGAAIITASAELIVTPPYSDKSQNTVTRISLE	718		
QY	863	GDIGNAY-----	-DKGFTNDTKIGAGVGVWRMASPVGQVREVDATGV	902	
Db	719	VDASAVWMTKMKSDKNGLESVDLKRLPDYGKSSIRASTGVGQWQSPIGLPLVSTAKPI	778		
QY	903	K--EEGNPIKILHFIQTFP	919		
Db	779	KYENDVEQFQFISIGSF	797		
RESULT 38					
AAR53754					
ID AAR53754 standard; Protein: 797 AA.					
XX AAR53754;					
AC					

DT	08-DEC-1994 (first entry)
XX	
DE	H. influenzae b Ca D15 sequence.

KM Vaccine; passive immunization; vector; antiserum; diagnosis; D15;  
KM OMP; outer membrane protein; H1b.

OS Haemophilus influenzae type b ca strain.

PN WO9412641-A.

PD 09-JUN-1994.

PF 23-NOV-1993; 93WO-CA00501.

PR 23-NOV-1992; 92GB-0024584.

PA (CONN-) CONNAUGHT LAB LTD.

PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;

XX  
11

DR WPL; 1994-200203/  
DR N-PSDB; AA066198.

DR N-PSDB; AAQ66198.

aa Nucleic acid encoding D15 outer membrane protein - esp. of  
PT

PT etc. useful in vaccines, for diagnosis and for passive

PT immunisation.

PS Disclosure; Fig. 1A; 161pp; English.

Outer membrane protein (OMP). D15 genes were isolated by screening chromosomal libraries of *H. influenzae* type b (Hb) strains Ca, Eagan and Minn A, and the non-typable (NTHI) strains 5833 and PAK 12085. Nucleotide sequences were determined for the D15 genes (AA065198-202) and the corresponding sequences were derived (AA853754-56). D15 OMP can be produced easily and on a large scale free of other antigens and lipooligosaccharides, by recombinant DNA methods using the isolated genes.

Sequence 797 AA;

Query Match 3.48; Score 161.5; DB 15; Length 797;

Best Local Similarity 19.58; Pred. No. 0.017;

[illegible]

DB 409 VDVYKVK-----ERNMGSIINGIGYTESGISYQASVKQDNFLGTGAASVAG 457  
QY 609 NNDGQACAEFLSE---DKGKYLATKPLSHPLNDQLRATLGYQOEYFGHS-TNGFDL 664  
DB 458 TRNDYGTSVNLGTYTEPYTFKDGVSLOGNVFPENYDMSKSDTSSNKRRTYGSVTLGFPV 517  
QY 665 STRT-----LEH---EISRTIIONGCMWRT---YSLRYRLDKLKTQAPPEMDELVDV 713  
DB 518 MNNSIYYVGLGHTYKISNFALF---YNNRLYIQSMKFKNGIKTN-----DFDS 565  
QY 714 NGPSEQALLACVAHVKTADNLVNPMPRGYRORYSLEVSSGLV---SDANMALIARAGI 769  
DB 566 FGM-----NYSNLMRGYFPFKGKASLGWHTIPGSDNKKYKRLSADY 607  
QY 770 SCVYSGDNAYGNSRAH--QMTGGIOAGYIWSDNFNHVPYRLRFPGAGDOSIRGVAHDSL 827  
DB 608 QOEYPL-----DRDHLWVSAKASAGANGFNKRRLPFYQTYTAGGIGSLGFPYGS 660  
QY 828 SP-----ISDKGLTGGQVLAAGTAE--YNEPMKD-----LRAYF 862  
DB 661 GPMATYAEYNGSGTGTFFKISSD--VIGGNALATASPELIYPTPFVSDKSQNTVATSLF 718  
QY 863 GDIGNAY-----DKGFTNDTKIGAGVGRMASPVGYRVDAVATGV 902  
DB 719 VDAASVMMTKMKSDKNKGLSDVLRPLDYKSSRIASTGVGFQMSPIGLIVESYAKPI 718  
QY 903 K--EEGNPIKLAFFIGTTP 919  
DB 779 KYENDVDEOFQFSTIGSF 797

RESULT 39  
AAR45002  
ID AAR45002 standard: Protein: 1319 AA.

AC AAR45002;

DT 13-JUN-1994 (first entry)

DE Cellulose synthase operon, gene C product.

KM Bacterial: cellulose synthase: operon; gene A: gene B: gene C;  
KW gene D: transcription vector; recombinant microorganism;  
KW cellulose synthesis.

XX Acetobacter xylinum.

XX US5268274-A.

XX 07-DEC-1993.

PF 12-APR-1989; 89US-0337194.

PR 10-APR-1990; 90CA-2014264.

PR 11-APR-1990; 90IE-0001317.

PR 09-APR-1990; 90IL-0094053.

PR 12-APR-1990; 90NZ-0233312.

PR 04-APR-1990; 90WO-US01811.

XX (CETU ) CETUS CORP.

XX Ben-Bassat A, Benziman M, Calhoun RD, Fear AL, Gelfand DH;

PI Meade JH, Tai R, Wong H;

XX WPI: 1993-404004/50.

DR N-PSDB: AAO53522.

XX Contiguous nucleic acid sequences - encoding bacterial cellulose  
PT synthase  
XX Claim 52; Fig 1 and Columns 53-72; 79pp; English.  
XX The sequences given in AAR45000-03 represent the proteins encoded by  
CC

CC the bacterial cellulose synthase operon. The sequence given in  
CC AAR45004 is the beginning of an open reading frame overlapping the end  
CC of this operon. The bacterial cellulose synthase operon contains  
CC four genes, genes A-D. The operon sequence may be used in a  
CC transcription vector for the expression of the cellulose synthase  
CC operon to increase cellulose production in a recombinant microorganism.  
CC This system may be used as an important tool for exploring mechanisms  
CC of cellulose synthesis and for enhancing production of cellulose.

XX Sequence 1319 AA:

Query Match 3.4%; Score 159.5; DB 14; Length 1319;

Best Local Similarity 19.2%; Pred. No. 0.047;

Matches 201; Conservative 131; Mismatches 360; Indels 357; Gaps 51;

QY 43 PAHDIAI-----NQAKAGNPVLLT-PEQIQARL--NAAGINAKPQSGAL- 84  
DB 164 PPHSLAVEYYQTWAGYPAQDOARAGIAGVAVSNPDYRQALFPAQALTYNTSTRMGLT 223  
QY 85 ---DVNFPDOQSPI-SRIGOSPPLGID-MSVIEETPLSLEELFAQESTEMGINPDYI 139  
DB 224 RLKDIQSFSQAEVEAAQAQSTROTLSWLPVNPETQPLMEQLSA-----HPMD-- 273  
QY 140 PEYQGEQPNSEVVVPTLEPEKPLIKRLYARLFNDGVNKPRLKAKFYQSOSGETSAI 199  
DB 274 ---TALREHMLHPGGPPDKAGIARQ-----AGYQQLNAGRILAA 310  
QY 200 GSSHQKTEPYANIKAALEDITQESAMDNGSIFRLROTALVAARANGYDIDSLINSTI 239  
DB 311 EDSFQ-----SALQINS-----HDAD-----SL 328  
QY 260 GEVDVYIHDGEPVYIDYRAVEYRGEGADPKAFTVADEVPFLIGVFHHKYEKTKNLI 319  
DB 329 GGMGLVSMKQGDYA--EARKFEEMAADPK-----TADRRKRALGAVSGETASVROLI 362  
QY 320 ENASAEHGYFFDGRW-----LDRSYDVI----- 341  
DB 383 ---AAHQYTEAKQOLATLARQPGQYTGATLMLADLQIRSTGQIAAAEQEYRGILSREPNN 438  
QY 342 -----LPDNTADV-SLIYDTGYOY--RPFDEVYFTIDPKTNQLTTPDKLPVK 366  
DB 439 QIALMGILARYDMAQNTAEKROLLSRGPQYASQVEIEVSGLMAAASQTSARAKVSIL 498  
QY 387 RELLEQL-----LTVNMGEYVNLQ-----AVRAL-----SND 413  
DB 499 REAMAQAPRDPWYRINLANALQOQGDVAEGRWQPILANPYARQDRQAGIILTYGSGND 558  
QY 414 LIAIRYFNMVNTETIVEPERKQIQNDQVSFEQSSSKTEPAQVDESTLEPIETVELTDGI 473  
DB 559 AMTRQLLAGLSPADYSPAIRSIA-EEMETIKODLASRLSWSNPVLIREALTQPPDTPGAR 617  
QY 474 LMDISPIEFSASNLIODKLNLVAKARHLVDMDDVY-----LAINHDGCVARSITGRI 527  
DB 618 GVAVADLEFRQGDVHARMAALRIASTRTTIDLSPODRISTATEYMKISNPVAAAR-LIAPL 676  
QY 528 SPAV-SAVARAIIIPDESE-----NEVIDLPERFALNRKTPADVYOSKPYLVFVASDK 581  
DB 677 GDTGTSATGSAIILPEQVYQLQQLRMGISVAQSDLLNORGQQAQAYHILAPALQADEATS 736  
QY 582 PR-----DQIGLIGMSDGTGRLVTKFEHNLINRDGYQAGELRUSDEKGV-KL 630  
DB 737 PKIALARLYNGHKPKGALEID-----LAVLRHNPQDLARQAAYQAAVSDHNSLATRL 791  
QY 631 YATKPLSHPLNLQRLATLGYQOEYFGHSTNGPFLSRTLEHLSRIIONGCMNRTYSLR 630  
DB 792 AMDGVQESPMADARAWLAVADQDGH-----GORTLE-----DLRRAYDLR 833  
QY 691 YRL-----DKLTKQAPPET-----WQDLPVDFVNGKPSQEA----- 721  
DB 834 LQOVBECTRAASGAGAQDEAL---APPSTNPPFRPRRGYHQTETELGAVTGGSYAEAAASP 890  
QY 722 ---LLAGVAHV-KTVADNL---VNPMPRGYRORYSLEVSSSGILVSDANMAIA-----RAG 768  
DB 722 ---LLAGVAHV-KTVADNL---VNPMPRGYRORYSLEVSSSGILVSDANMAIA-----RAG 768



DB 891 TSDOMLSIAQITLTENLAPSIDGLGFRSR-SGEHG-MGRLEANIPIVGRPLQAG 948  
QY 769 ISGVYSPGDNAVGSNRAHOMTGTIOAGYIWSDNFN-----HVPYRLRFAGDQDSIRGY- 822  
DB 949 ASA-----LFSITPTMIMSGNLTGSIYVFP---RY--CTMMGVQAIN 987  
QY 823 AHSLSLPSIDKGYITGG---QVLAVGTAEYVEFMKDLRLAVFGDIGNAYDKGFTNDTKI 879  
DB 988 QYDS-----YTNACRDQDRIAGTAE-----AGFAPDVQF 1017  
QY 880 G-----AGVGRMASPYGVRYDVATGVK 903  
DB 1018 GNSWVRADVC---ASPIGFPITNVLGVE 1043  
RESULT 40  
ABG21666  
ID ABG21666 standard; Protein; 1249 AA.  
XX  
AC ABG21666;  
XX  
XX 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #21657.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
XX N-PSDB; AAS85853.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID NO 52025; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 1249 AA;  
Query Match 3.38; Score 157.5; DB 22; Length 1249;  
Best Local Similarity 18.28; Pred. No. 0.06;  
Matches 114; Conservative 88; Mismatches 181; Indels 245; Gaps 27;  
QY 466 TVELDGLTMDISLPSASNLIDKLNLA----- 496  
DB 54 TITVINGVLDLVNDQDFTEAFRTDKLELTSGNIADHNGVYSGVFDIHSSDYLMADLV 113  
QY 497 -----AKARHLYD---MPDDRVLAINH---DDGVN-----RSILG-----RISDA 530  
DB 114 NDRRTDTSKSNYGYGIYAMNSDGLHTTINGNDVDNGTMIOLCNVVAATGNKYVRIDNA 173  
QY 531 VSAVARAILPDESENEVIDPE--RTALANKRPADV---YQSK----- 570  
DB 174 TGAGANA---DYKDKETIYVDVNSNATFSANKRADLGAYTYQAEQGRNTACLPQMELEDD 230  
QY 571 ---VPLVYFVASDKPRDQIGLGW--GSDT--GTRLYTKFEHNLINRDCYQAGAELRUSED 624  
DB 231 YANPPLSIPSANTNI-----WNLGQDVTGTRL-TNSRHGLADNGS----- 269  
QY 625 KKGVLVATKPLSHPLNDQLRATLGYOEYVGHSTNGFDSLTRLERHSRIT-----QN 680  
DB 270 -----AWYSY-----FGNPNCG-DNCTINYDDVNGIMVGVDTKI 303  
QY 681 GGMNRTYSLRYRLDKLTKTOAPPETMODLPVD-----FVNGKPSQEAL---LAGVAVHKT 731  
DB 304 DGNNAKWIVGAAGSLKLVDDQSFATYIYSSAHFANNVFVDSLSLSYHFNNDLSATMSMGT 363  
QY 732 VADNLVPMR---GYRQRTSLENGSSGLVSDANMAIRAGISGYSPGDNAVGSN----- 783  
DB 364 YVDGTSNDAMWFGKLGAGYFKLGADAGYVT-----PYGSVGLFQSGDDYQLSNDKQVD 417  
QY 784 -----RAHOMT-----GGIAGCYIMSDNFH 804  
DB 418 QGSYDSMRYESPRVEPASACKMKVTLDTATYPIIDDDHKWVYLGRWYGDGLGKE 477  
QY 805 VPYRLRFAGGQDSIRGYANDSLSPIS-----DKGY-----L 836  
DB 478 MPFEYENFYAGGSSYVRGFSQNTIGPKAVYFPHQASNTDPDVEDYCATQDGAJDKCKSDDA 537  
QY 837 TGGQVLAVGTAEY-----NYEPMKDLRLAVFGDIGNAYDKGF-----TND 876  
DB 538 VGGNALAVASLEFITPPFISDKYANSVRTSFEWDMCTVMDTNWSSQYSGYPDYSDPSN 597  
QY 877 TKIGAGVGRMASPYGVGVRYDVATGVKE 904  
DB 598 IRMSAGIALQMSPLGLPVSFAOPFRK 625

Search completed: April 28, 2003, 16:21:39  
Job time : 101 secs





```
Db 72 DVKAHQEGDVLVSVAKSIISDVKIKGNSVITPEALKQNDANGKGVDL-----IREK 127
Qy 316 KNLENSAHEGHYFDRMLDRSDVY---LPDNTADVSLYDTGTQYREDEVFETIDPK 372
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 LNEFAKSVKEHYASVGR-NATVEPIVNTLPNNRAELIQINEDDKAKLASLTF-----K 181
Qy 373 TNQITDPPKLPVKRELEQLLVNMGAY-----NLQAVRALSNDLITRFYNNVNT 425
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 GNE-SVSSSTLOEQMELQPSMWKLMKNKFEGAOFEKDQAIR-----DYLLNGYA 232
Qy 426 EIVPEPEQIQNDQVSESSSSRTPEAOVD-----ESTLEPIETVELTDG 472
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 233 KAQITKTQVQNDKTKVNTITDVNEGLQYDLRSARIIGNLGMSELEPILLSLHLNDT 292
Qy 473 I-LMDISPIEFASNLQDKLNLVAKARHLVMPD---DRVLAINHDDGVNRSIL--- 524
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 FRRSDIADVE---NAIKALGERGYGNTTVNSVPDPAKNTLAITFVDAGRLLTVRQ 348
Qy 525 ---GRISDAVSAVARAILPDES---ENEVIDLP-----ALANKRTPAD----- 564
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 349 LRFEGNTVSA DSTLRQEMNQEGTWNLSQVELGKIRLDRGTGFEFTVENRIDPLNGSND 408
Qy 565 ---VYQSKVPLVYVVASDKPRDQI--GLMGSDTGTRLVTKF-EHNL----- 608
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 409 VDVYVKV-----ERNTGSINFGIGYTESGISYQTSIKQDNFLGTGAANSIAG 457
Qy 609 NRDGVOAGAEIRLSE---DKKVKLYATKPLSHPLNDQLRATLGYOQEVFGHS-TNGFDL 664
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 458 TKNDGTSVNLGTYTEPYFTKDGVSIGGNIFFENYDNKSDTSSNKRRTYGSNVTLGFPV 517
Qy 665 STRT-----LEH---EISRSIIQNGCMNRT---YSLRYRLDKLTKQAPETMODLVDEV 713
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 518 NENNSYVGLGHTYKIKISFALE---YRNRLYIQSMKFKGNGKITN-----DFDS 565
Qy 714 NCRPSQEQALLAGVAVHKVYADNLVNPMRGVRQRYSLVGVSSGLV---SDANMAIARAGI 769
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 566 FGW-----NYNSLNRGYFPTKGYKASLGGRVITPGSDNKKYKLSADV 607
Qy 770 SGVYSFGDNAYGNSRAHO--MTGTOAGYIMSDNHNHVPYRLRFAGDQSIIRGYAHDLS 827
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 608 QGFYPL-----DRDHMVVYSAKASAGYANGFGNKRLPFYQTYAGIGSLRGFAGYGI 660
Qy 828 SP---ISDKGYLT-----GGQVLAVGTAE--YNEEFKMD-----LRLAVFEDIGNA 868
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 661 GPNALYAEHNGCTENKISSDVIIGNAITTAABELIVPPFVSDKSQNTVRSLSFVDASV 720
Qy 869 Y-----DKGFTNDTKIGAGVYMASPVGVQGVADVATGVK--REG 906
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 721 WNTKMSDKNGLSKVYLDLPDYGKSSRIASTGVGFQMQSPFIPGVFSYAKPIKYEEND 780
Qy 907 NPIKLHFFICTPF 919
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 781 DVEQFPQFSGCF 793

RESULT 2
US-09-135-166-10
: Sequence 10. Application US/09135166
: Patent No. 6083743
:
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Slim & McBurney
: STREET: 67H Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
```

```
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/135,166
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/433,522
FILING DATE: 12-SEP-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-135-166-10
```

```
Query Match 3.8%; Score 180.5; DB 3; Length 793;
Best local Similarity 20.1%; Pred. No. 1,7e-06;
Matches 159; Conservative 121; Mismatches 306; Indels 207; Gaps 39;
```

```
Qy 263 DVIITHDLGEVYIDYRA-----VEVRGEQA-DDKAFYTADEVPRLIGVFNHGKETK 315
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 72 DVKAHQEGDVLVSVAKSIISDVKIKGNSVITPEALKQNDANGKGVDL-----IREK 127
Qy 316 KNLENSAHEGHYFDRMLDRSDVY---LPDNTADVSLYDTGTQYREDEVFETIDPK 372
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 LNEFAKSVKEHYASVGR-NATVEPIVNTLPNNRAELIQINEDDKAKLASLTF-----K 181
Qy 373 TNQITDPPKLPVKRELEQLLVNMGAY-----NLQAVRALSNDLITRFYNNVNT 425
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 GNE-SVSSSTLOEQMELQPSMWKLMKNKFEGAOFEKDQAIR-----DYLLNGYA 232
Qy 426 EIVPEPEQIQNDQVSESSSSRTPEAOVD-----ESTLEPIETVELTDG 472
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 233 KAQITKTQVQNDKTKVNTITDVNEGLQYDLRSARIIGNLGMSELEPILLSLHLNDT 292
Qy 473 I-LMDISPIEFASNLQDKLNLVAKARHLVMPD---DRVLAINHDDGVNRSIL--- 524
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 FRRSDIADVE---NAIKALGERGYGNTTVNSVPDPAKNTLAITFVDAGRLLTVRQ 348
Qy 525 ---GRISDAVSAVARAILPDES---ENEVIDLP-----ALANKRTPAD----- 564
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 349 LRFEGNTVSA DSTLRQEMNQEGTWNLSQVELGKIRLDRGTGFEFTVENRIDPLNGSND 408
Qy 565 ---VYQSKVPLVYVVASDKPRDQI--GLMGSDTGTRLVTKF-EHNL----- 608
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 409 VDVYVKV-----ERNTGSINFGIGYTESGISYQTSIKQDNFLGTGAANSIAG 457
Qy 609 NRDGVOAGAEIRLSE---DKKVKLYATKPLSHPLNDQLRATLGYOQEVFGHS-TNGFDL 664
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 458 TKNDGTSVNLGTYTEPYFTKDGVSIGGNIFFENYDNKSDTSSNKRRTYGSNVTLGFPV 517
Qy 665 STRT-----LEH---EISRSIIQNGCMNRT---YSLRYRLDKLTKQAPETMODLVDEV 713
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 518 NENNSYVGLGHTYKIKISFALE---YRNRLYIQSMKFKGNGKITN-----DFDS 565
Qy 714 NCRPSQEQALLAGVAVHKVYADNLVNPMRGVRQRYSLVGVSSGLV---SDANMAIARAGI 769
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 566 FGW-----NYNSLNRGYFPTKGYKASLGGRVITPGSDNKKYKLSADV 607
```



ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/433,522A  
FILING DATE: 12-SEP-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I  
REGISTRATION NUMBER: 24,973  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 792 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-433-522A-8

Query Match 3.8%; Score 178; DB 3; Length 792;

Best Local Similarity 20.2%; Pred. No. 2.8e-06;  
Matches 160; Conservative 118; Mismatches 306; Indels 210; Gaps 39;

QY 263 DVIHIDGEPYIDYRA-----VEVRGEG-ADDKAFTTVADDEVPLLIGDVFHNGKYEK 315  
DB 72 DVKAHOGEDVLVSVAKSISIDVKIKGNSIIPPEALKQNDANGFKYGDIL-----IREK 127  
QY 316 KNIENASAEHGIDGKRLDRSVYI---LPDNTADVSLIYDTGTQYRFDEVEVFTIDPK 372  
DB 128 LNEFAQSVKEHYASVGRY-NATVEPIVTLTPNNRAEILIQINEDDKAKLASLTF-----K 181  
QY 373 TNOITTPDKLPVARELLEQLLTVMNGEAY-----NLQAVRALNSLDLTRFENMYNT 425  
DB 182 GNE-SVSSSTLQDEMELQPODSMMKLMKKEGQAFERKDLQAIR-----DYLNNGYA 232  
QY 426 EIVPEREQIONDVSEQSSSSRTERPAQVD-----ESTLEPYIETVELTDG 472  
DB 233 KAQITKADVOLNDEKTKVNTIIVNEGLOYDRLRSARITLNGMSAELEPLSLALHNDT 292  
QY 473 I-LMDISPIEFSSNLQDKLNLVAARHLIYDMPD---DRLALNHDDQVNSIL--- 524  
DB 293 FRNSDIADVE---NAIKAKLGERGYNTTVNSVDPEDDANKTLAIFVVDAGRRLTVHQ 348  
QY 525 -----GRISDAVSAAVARAILPDES---ENEVIDLP-----ERT-----ALANKRTPAD----- 564  
DB 349 LRFEGNVSADSTLRQMRQOEGTWYNSQLVELDKIRLDRIGFEFEYENRIDLPIINGSND 408  
QY 565 ----YQSKKVPPLVFEVASDKPRDQOI---GLGWSGDTGTRLVTKF-EHNLI----- 608  
DB 409 VDVYVYKVK-----ERNTGSINFGIGYGTESISYQASVKQDNFLCTGAASVAG 457  
QY 609 NRDOYGAGAEILRSE---DKRGVLYATKPLSHPLNDQLRATLGLQYQGEVGH-S-TNGD 664  
DB 458 TKNDYGSIVLGLTEPFTFDGVSGLGNVFPENDNSKSDTSYKRTYGSNTVLGFPV 517  
QY 665 STRT-----LEH---EISRSIIIONGMNRT---YSLRYRLDKLTKQAPEPTWODLPVDFV 713  
DB 518 NENNSYVVGCHTYNKSINALE---YNRNLIYQSMKFKGIGITN-----DDFS 565  
QY 714 NGRPSOALLAGVAHVKTADVNLVNPGRYRORYSLEVGSGLV---SDANMAIARAGI 769  
DB 566 FGV-----NYNSLNRGYFPTKGVKASLGGRYTITPGSNKYYKLADY 607  
QY 770 SCYYSFGDNAGYNSRAH---QMTGGIAGYIMSDNFNHPVRLRFFAGDDOSIRYAHDSL 827  
DB 608 QGTYPL-----DRDHLVNVYSAKASAGVANGFNKRRLPYQYTYTAGGIGSLRGFAVAGSI 660

QY 828 SP-----ISDKGYLTGGOVLAVGTAE--YNTEFMKD-----LRLAVEGDIGN 867  
DB 661 GPNALYOGONNKNKNISSD--VIGGNAIATASAEILVPTPFVSDKSQNTVRSLSFVDAAS 718  
QY 868 AV-----DKGFTNDTKIGAGYGVWASPYGOVRVDVATGVK--FE 905  
DB 719 VVNTYKWSDKNGLESNYLKDLPDYGRKSRTRASTGVGFOWQSPSGFVPSYAPIKKYEN 778  
QY 906 GNPRLHFEIGTFP 919  
DB 779 DVEQFOFSGISGF 792

## RESULT 5

US-09-135-166-8  
Sequence 8, Application US/09135166  
Patent No. 6083743

## GENERAL INFORMATION:

APPLICANT: CHONG, Pele  
APPLICANT: THOMAS, Wayne  
APPLICANT: YANG, Yan Ping  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: SIA, Dwo Yuan Charles  
APPLICANT: KLEIN, Michel  
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: 6TH Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada

ZIP: M5G 1R7  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/135,166  
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/433,522  
FILING DATE: 12-SEP-1995

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I

REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:

LENGTH: 792 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-09-135-166-8

Query Match 3.8%; Score 178; DB 3; Length 792;

Best Local Similarity 20.2%; Pred. No. 2.8e-06;  
Matches 160; Conservative 118; Mismatches 306; Indels 210; Gaps 39;

QY 263 DVIHIDGEPYIDYRA-----VEVRGEG-ADDKAFTTVADDEVPLLIGDVFHNGKYEK 315  
DB 72 DVKAHOGEDVLVSVAKSISIDVKIKGNSIIPPEALKQNDANGFKYGDIL-----IREK 127  
QY 316 KNIENASAEHGIDGKRLDRSVYI---LPDNTADVSLIYDTGTQYRFDEVEVFTIDPK 372  
DB 128 LNEFAQSVKEHYASVGRY-NATVEPIVTLTPNNRAEILIQINEDDKAKLASLTF-----K 181



Db 661 GPMALYOGNNKFNKISSD--VIGGNALATASABELIYPTPEVSDKSONTQVTSLEFVAAS 718  
Qy 868 AY-----DKGFTNDPKIGAGVGNRNASPVGOVVDVATVVK--EE 905  
Db 719 VWNTRKMSDKNGLESANLKLDPDYGKSSRRIRASTGVGFQWQSPGPPVYFSAKPIKRYEN 778  
Qy 906 GNPDKLHFFIGTPE 919  
Db 779 DDVEQFQFSGISGSF 792

RESULT 7  
US-08-433-522A-2  
Sequence 2, Application US/08433522A  
Patent No. 6013514  
GENERAL INFORMATION:  
APPLICANT: CHONG, Pele  
APPLICANT: THOMAS, Wayne  
APPLICANT: YANG, Yan Ping  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: SIA, Dwo Yuan Charles  
APPLICANT: KLEIN, Michel  
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN  
NUMBER OF SEQUENCES: 55  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: 6TH Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/433,522A  
FILING DATE: 12-SEP-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: STEWART, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 797 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-433-522A-2

Query Match 3.6%; Score 171.5; DB 3; Length 797;  
Best Local Similarity 19.8%; Pred. No. 1e-05;  
Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

Qy 263 DVIHDLGEPVYIDYRA-----VEVREGGA-DDKAFYTVADEVPLLIGVFHNGKYEK 315  
Db 72 DVKAHQEGDLVAVYAKSIISDVKIKNGSVIPTKALKNLDANGFVGVVL----IREK 127  
Qy 316 KNIENASAEHGCFGDRWMLDRSDVI---LPDNTADVSLIYGTGYRFDVVFETIDPK 372  
Db 128 LNEFAKSVKEHYASVGRY-NATVEPIVNTLPNNRAEILIOINEDKAKLASLTF-----K 181  
Qy 373 TNOITDPPKLPKRELLLEQLTVNNGEAY-----NLQAVRALSNDLATRYFNKVVNT 425  
Db 182 GNE-SVSSSTLQOMELQDPDSMMKLMKNKEGAOFEKDLOSIR-----DYLLNNGYA 232  
Qy 426 EIVFPEREIONDQVSFEQSSSSSRTEPAQVD-----ESTLEPIETVELTLDG 472

Db 233 KAQITKTDVQINDEKTKVNVITDVNEGLQYDLRSARITIGNLGMSELEPLISALHNDT 292  
Qy 473 I-LMDISPIERSASNLIDDKLNLVAKARHLXDMPD-----DRVLINDDGVNRSLT--- 524  
Db 293 FRRSDIAYE-----NATAKIGERGYSATVNSVDFPDANKTLAITLVDAGRRLYRQ 348  
Qy 525 ---GRISDAVSAVARAILPDES---ENEVIDP---ERT---ALANRKT PAD----- 564  
Db 349 LRFEGNTVSAADSTLHQERROEGTWNSQLVELGKIRIDRTGFEFTVENRIDPINGSND 408  
Qy 565 ---YYQSKVPLVYFVADSKPRDGI--GLGWSPTGTRLYTKF--EHLI----- 608  
Db 409 VDVIYKYK-----ERNYGSINFGIGYTESGISYQSVKQDNFLTGAAVSIAG 457  
Qy 609 NRDGQAGAEIRLSE---DKKGVKLYATKPSHPPLNDQLRATIGYQOQVFGHS-TNGFDL 664  
Db 458 TKNDYGTSVNLGYTPRYTTKGVSIGGNVFPENTYNSKSDSSNKRTTYGSAVTVLGEFPV 517  
Qy 665 STRT-----LEH---EISRSIIQNGWNR---YSLRYRLDKLTKTQADPETWQDLPVDFV 713  
Db 518 NENNSYVYGLGHTYKNKISNFALF---YNRNLTYIQMKFKGNGIKTN-----DFDFS 565  
Qy 714 NGKPSQELLAGVAVHKTVADNLVNPMPRGYRQRYSLFVSGSLV-----SDANMALARGI 769  
Db 566 FGW-----NYNSLNRGYPTPKGVKSLGGRVITPQSDNKYKYLSDV 607  
Qy 770 SGVYSGFDNAYGNSRAH--QMTGGIAGYIWSDNENHPYRLRFPGAGDQIRGYAHS 827  
Db 608 GGFYPL-----DRDHLVYVSAKASAGYANGFGKRLPFQTYIAGSIGSLRGVAFSI 660  
Qy 828 SP-----ISDKGYLTGQYLVAGTAE--YNEFMKD-----LRVAF 862  
Db 661 GPMALYAEYNGSGTGTFPKKISSD--VIGGNALATASABELIYPTPEVSDKSONTQVTSLEF 718  
Qy 863 GIDINAY-----DKGFTNDPKIGAGVGNRNASPVGOVVDVATVVK 902  
Db 719 VDASVWNTKRSKDKNGLESVDLKLRLPDYGKSSRRIRASTGVGFQWQSPGPPVYFSAKPI 778  
Qy 903 K-BEGNPDKLHFFIGTPE 919  
Db 779 KKYENDVEQFQFSGISGSF 792

RESULT 8  
US-08-433-522A-4  
Sequence 4, Application US/08433522A  
Patent No. 6013514  
GENERAL INFORMATION:  
APPLICANT: CHONG, Pele  
APPLICANT: THOMAS, Wayne  
APPLICANT: YANG, Yan Ping  
APPLICANT: LOOSMORE, Sheena  
APPLICANT: SIA, Dwo Yuan Charles  
APPLICANT: KLEIN, Michel  
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN  
NUMBER OF SEQUENCES: 55  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: 6TH Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/433,522A  
FILING DATE: 12-SEP-1995  
CLASSIFICATION: 435



ATTORNEY/AGENT INFORMATION:  
 NAME: STEWART, Michael I  
 REGISTRATION NUMBER: 24,973  
 REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (416) 595-1163  
 TELEFAX: (416) 595-1163  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 797 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-433-522A-4

Query Match 3.6%; Score 171.5; DB 3; Length 797;  
 Best Local Similarity 19.8%; Pred. No. 1e-05;  
 Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

QY 263 DVIHIDGEPYIDYRA-----VEVRGEA-DDKAFITVADEVPLLIGVFHHGKYETK 315  
 DB 72 DYKHOEGDVLVSVVAKSIISDVKIKGNSVIPTKALQNDANGKRVGDV-----IREK 127  
 QY 316 KNLENSAEGHYDGRWLDKRSVDI---LPDNTADVSLIYDTGTQYRFDEVVFEFTIDPK 372  
 DB 128 LNEPAKSVKHHYASVGR-NATVEPIVNTLPNNRAEILIQINEDDKAKLASLTF-----K 181  
 QY 373 TNOITTDPKLPVKRELLLEQLLTVMGEAY-----NLQAVRALSDLLATRYFNNVNT 425  
 DB 182 GNE-SVSSSTLQEQMELQPSWMKMGKNGKFEQAQFEKDLOSIR-----DYLLNNGYA 232  
 QY 426 EIVPEPERQIQNDQVSEQSSSRTEPAQVD-----ESTLEPIETVELTDG 472  
 DB 233 KAQITKTQVQLNDEKTKVNTYIDVNEGLOXDLSARILIGNLGSMALEPLLSALHLNDT 292  
 QY 473 I-LMDISPIEFASNLQDNLNVAAKARHLVMDP-----DRVLAINHDDGVNSIL--- 524  
 DB 293 FRNSDIADVE-----NAKKAALGERGYSATVNSVPDDANKLATILVVDAGRRLLVRQ 348  
 QY 525 ---GRISDAVSAVARAILDPDES---ENEVIDLP-----ALANKRTPAD----- 564  
 DB 349 LREGMTVSAADSLRQEMROEGWTWYNSQLVELKILDTGFEFEYENRIDLPIGNSNDE 408  
 QY 565 ---VYQSKVPLVYVAAKPRDQGI---GLGMSDGTGRLVTKR-EHNLI----- 608  
 DB 409 VDVVYKVK-----ERNMGSGINFGIGYEGSGISYQASVKQDNPLGTGAASVAG 457  
 QY 609 NRDGYQAGAEIRLSE---DKKGYKLTATKPLSLPLNDQLATLGYQOEVEFGHS-TNGFDL 664  
 DB 458 TKNDYGTSVNLGYTEPYFTKDGVSLSGNVFEFENYDMSKSDTSSNYKRTYGSNVTLGFPV 517  
 QY 665 STRT-----LEH---EISRSIIIONGGWNR-----YSLRYRLDKLTKQAPEPTWQDLPPYDFV 713  
 DB 518 NENNSYVYGLGHYNNKISNALE---YNRLYIQSMKFKKNGIKTN-----DFDFS 565  
 QY 714 NGKPSOEALLAGVAHKTVAADNLVNPGRYRORYSLEVGSSGLV---SDANNAIARAGI 769  
 DB 566 FGV-----NYNSLNRGYFPFKGYKASLGGVYTIIPGSDKNKTKLSADY 607  
 QY 770 SCVYSGCDNAAGSNAH---QMTGGIQAQIYWSNENHVPRLRFPKAGGDSIRNGYAHDSL 827  
 DB 608 QGTYPL-----DRDHLWVYSAKASAGYANGFGNKRRLPEYQYTAGGISLGLGFAVGI 660  
 QY 828 SP-----ISDKGYLTGGQVLAVGTAE---VYVEPKKD-----LRALVF 862  
 DB 661 GGNATIAEYNGSGTGTFKKIISD---VIGONMAIATASAEILVTPPVYSDKSNQNTVTSLSF 718  
 QY 863 GDIGNAY-----DKGFTNDTKIGAGVGVYRWMAAPVQGVQVNDVATGV 902  
 DB 719 VDAASVWMTKMKSDKNGLESVDLKLRLPDYGGKSSRIASVGVGQWQSPICQPLVFSYAKPI 778  
 QY 903 K---EEGNPTKLHFFIGTTPF 919

DB 779 KKYENDVEQFQFSIGSSE 797

RESULT 9  
 US-08-433-522A-6  
 Sequence 6, Application US/08433522A  
 Patent No. 6013514  
 GENERAL INFORMATION:  
 APPLICANT: CHONG, Pele  
 APPLICANT: THOMAS, Wayne  
 APPLICANT: YANG, Yan Ping  
 APPLICANT: LOOSMORE, Sheena  
 APPLICANT: SIA, Dwo Yuan Charles  
 APPLICANT: KLEIN, Michel  
 TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN  
 NUMBER OF SEQUENCES: 55  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sim & Mcburney  
 STREET: 6TH Floor, 330 University Avenue  
 CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R7  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/433,522A  
 FILING DATE: 12-SEP-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: STEWART, Michael I  
 REGISTRATION NUMBER: 24,973  
 REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (416) 595-1163  
 TELEFAX: (416) 595-1163  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 797 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-433-522A-6

Query Match 3.6%; Score 171.5; DB 3; Length 797;  
 Best Local Similarity 19.8%; Pred. No. 1e-05;  
 Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

QY 263 DVIHIDGEPYIDYRA-----VEVRGEA-DDKAFITVADEVPLLIGVFHHGKYETK 315  
 DB 72 DYKHOEGDVLVSVVAKSIISDVKIKGNSVIPTKALQNDANGKRVGDV-----IREK 127  
 QY 316 KNLENSAEGHYDGRWLDKRSVDI---LPDNTADVSLIYDTGTQYRFDEVVFEFTIDPK 372  
 DB 128 LNEPAKSVKHHYASVGR-NATVEPIVNTLPNNRAEILIQINEDDKAKLASLTF-----K 181  
 QY 373 TNOITTDPKLPVKRELLLEQLLTVMGEAY-----NLQAVRALSDLLATRYFNNVNT 425  
 DB 182 GNE-SVSSSTLQEQMELQPSWMKMGKNGKFEQAQFEKDLOSIR-----DYLLNNGYA 232  
 QY 426 EIVPEPERQIQNDQVSEQSSSRTEPAQVD-----ESTLEPIETVELTDG 472  
 DB 233 KAQITKTQVQLNDEKTKVNTYIDVNEGLOXDLSARILIGNLGSMALEPLLSALHLNDT 292  
 QY 473 I-LMDISPIEFASNLQDNLNVAAKARHLVMDP-----DRVLAINHDDGVNSIL--- 524  
 DB 293 FRNSDIADVE-----NAKKAALGERGYSATVNSVPDDANKLATILVVDAGRRLLVRQ 348  
 QY 525 ---GRISDAVSAVARAILDPDES---ENEVIDLP-----ALANKRTPAD----- 564

```

Db      349  LRFEGTVASADSTLRFQMKGQSGTWYNSQVLKRLDRFGFEYENKRIDPIINGSNDE 408
Oy      565  ---VYOSKKAFLPVYVADSKPRDGOI--GLWGSDGTGLVTKP-ENHLI-----
Db      409  VDVVYVVK-----ERNTGSIINGIGIGTGESGISYOASVKONFLGTGAANSIAG 457
-Oy      609  NRDGYOAGAEFLRLSE---DKGVKLYATKPLSHPLNQLRATLTGLOEVEFGHS--TNGFDL 664
Db      458  TKNDVOTSVNLGTEPYFTFKDSVSLGCVNFFENDNSKSDSTSSVYKRTTGGSVNTLGFV 517
Oy      665  SFTT-----LEH---EISRSIIIONGMNRT---YSLRYRLDKLTQAPPEPTMODLPIVDFV 713
Db      518  NENNSYYVGLGHTYKKNISNFALE--YNRNLVIGSMKFKGICITN-----DEDFS 565
Oy      714  NCKPSQSEALLAGVAHKTVADNLVPMHGYQORISLEVGSSGLV----SPANNAIRACI 769
Db      566  FGM-----NYSNLNGYFFPTKGVASLGGRTTTPGSDKNKYKSLADV 607
-Oy      770  SGCVYFSGDNAYVGSNRAH--QMTGGIOAGYIWSDFNHPYRLRPFPGGDSIRGVADSL 827
Db      608  QGFYPL-----DRDLWVVSAAKASAGVANGFENKRLPLFYOTYTTAGIGSLNGFAYGSI 660
Oy      828  SP-----ISDKGYLTGGVLAAGTAE--VYEFMKD-----LRLAVF 862
Db      661  GPNALVAEYGNSGGTGFKKISSD--VIGNALATATASAEILVPTPVSDEKQMTVRTSLF 718
Oy      863  GDIGNAY-----DKGFTNDTRIGAGVGVBMASPVQGVARNVATGV 902
Db      719  VDASVWNTTKKSDKKNGLSEVYLKRLPDYKSKSRIRASTGVGFOMOSPIGLPIVSYAKPI 778
Oy      903  K--PEGNAIKLHFTGTPE 919
Db      779  KKYENDVEQFQFSGISGF 797

RESULT 10
US-09-135-166-2
Sequence 2, Application US/09135166
Patent No. 6083743
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: THOMAS, Wayne
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: SIA, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Slim & McBurney
STREET: 6TH Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/135,166
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/433,522
FILING DATE: 12-SEP-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-829 MIS.jb
TELECOMMUNICATION INFORMATION:

```

```

? TELEPHONE: (416) 595-1155
? TELEFAX: (416) 595-1163
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 797 amino acids
? TYPE: amino acids
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US-09-135-166-2

```

Query Match	3.6%;	Score 171.5;	DB 3;	Length 797;
Best Local Similarity	19.8%;	Pred. No. 1e-05;		
Matches 158;	Conservative 121;	Mismatches 305;	Indels 215;	Gaps 39;

QY	263	DVILHIDGEPYIYDRA-----	-VEVGEGA-DDKAFTVADEVPLLIGYFHHGKETK	315
Db	72	DYKAHQEGDVLVSVAKSIISDYKIKGNSVITPEALAKONLDANGKFGVDL----	-TREK	127
QY	316	KNLIENASAEHGYPDGRWLDRSVDVI----	-LPDNTADVLSLYDTQYRFPEVFEFTDPK	372
Db	128	LNFEPAKSVKEHYASVGR-NATVEPIYNTLPPNNAELLIOINEDKAKLASLTF-----	-K	101
QY	373	TNOLITTPDPLPKYKRELLLEOLLTYNNGEAT-----	-NLOAVRALSNDLIATRYFNKMYNT	425
Db	182	GNE-SVSSSTLQEOEMELQDPDSMWKLMKNKEFGAOFKEKDLOSIR-----	-DYLLNNGYA	232
QY	426	EIVPEREQIIONDOVSPEQSSSSRTPEAOVD-----	-ESTLEPIVETVELTQG	472
Db	233	KAQTKTDVQOLNEKTEFVNTITVNELOQYDLASARIQLNGMSALEPLLSALHLNPT		292
QY	473	I-LMDISPIEFSSASNLQDLKLNVAAKARHLYDMPD----	-DRVLAINHDDGVKRSIL--	524
Db	293	FRRSIDIAVDE-----NAIKAKLGERGYSATVNSVPDEDDANKTLAITLVVDAGRRLTVNQ		348
QY	525	-----GRISDVSVAVALILPDES-----ENEYIDLP-----EXT-----	-ALANKRTPAD-----	564
Db	349	LREPGNTVSAADSTLRQEMKROEGTMYNSOLVELGKIRLDRTGFEFTVENRIDIPIGNSDE		408
QY	565	---VYOSKKVPLVYVNASDKPRDGOI--GLGMSQDPTGRULYKF--EHNLI-----		608
Db	409	VDVVYKVK-----ERNMGISNFGIGYCTESGISYQASVKODNLTGGAIVSING		457
QY	609	NBRGYQAQAEPLRSE--DKKGVKLYATKPLSHPLNDOLRATLGYOQEEVFGHS-TNGEDL		664
Db	458	TKNDYGTSVNLGTEPYETFKDGSISLGANFEPENYDNSKSDTSSNKKRTYTGSNVTLGGPPV		517
QY	665	STRT-----LEH-----EISRSIIONGWMNT-----YSLRYLMDLKTQAPEPTMODLPVDFV		713
Db	518	NENNSYVVLGHTYKISNFALF-----YNNNLYTQSKEFGKNGINKTN-----DPFPS		565
QY	714	NGKPSQEALLAGVAVAKHTVADNLIYVPMRGRORYSILEVSSGLV-----SDANMATAIRAGI		769
Db	566	FGW-----NYSNLNGRYPFTTKGVKVASIGSGRYVITPSSDNKTYKRLSADV		607
QY	770	SGVYSPGDNAYNSNAH--QMTGGIQAQAGTINSDFNHWPYRLRFPAGGDQSIGYAHDSL		827
Db	608	QOEYPL-----DRDHLMVVSASAKASAGANGFGNKRLPFYOTYAGAGIGSLRPGYCSI		660
QY	828	SP-----ISDKGYLGGVGLANGTAE--YNEFEMD-----LRLAVF		862
Db	661	GPNAIYAEYGNSGIGTKFKKISSD--VIGGNAINATASALLIVPPPEVSDKSQNTVRSLSF		718
QY	863	GDIGNAY-----DKGFTNTDKIGAGYGVMSAEVQGOVRYVATGV		902
Db	719	VDASVYNTKWKKSXKNGLESVDLKRLPDYGKSSRLRSTAGVGFQWQSPISGLVYSYAKPI		778
QY	903	K--EEGNFIKHLHFTGTFP		919
Db	779	KKEYNDVEQFQFSTIGGSF		797

RESULT 11  
US-09-135-166-4



TELEFAX: (416) 595-1163  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 797 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-135-166-6

Query Match 3.6%; Score 171.5; DB 3; Length 797;  
 Best Local Similarity 19.8%; Pred. No. 1e-05;  
 Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

```

QY 263 DVIHDLGEVYIDYRA-----VEVRGEGA-DDKAFVTVADVPLLIGDVFHHGKYEK 315
DB 72 DYKAHQEGDVLVSVAKSIISDVKIKGNSVITPEALKQNLANGKRVGDL-----IREK 127
QY 316 KLIENASAEHGFCDGRMLDRSDVY---LPDNTADVSLIYDTGQYRDEVEVFTIDPK 372
DB 128 LNEFAKSVKEHVASGVR-NATVEPIVNTLPNNRAEILIOINEDDKAKLASLTF-----K 181
QY 373 TNGLTDPDKLPVKRLEQLTLVNNGEAY-----NLQAVRALSNDLIATRYFNKAVNT 425
DB 182 GNE-SVSSSTLQEQMELQDPSMMKLMGNKFEQAQFEKDLQSTR-----DYLLNGYA 232
QY 426 EIVFPEREQIQNDQVSEFQSSSRTEPAQVD-----ESTLEPIVETVELTDG 472
DB 233 KAQITFDVQUNDEKTKVNTIDVNEGLOYDLRSARITIGNLGMASLEPLLSALHLNDT 292
QY 473 T-LMDISPIEFASNLIDQKLNVAAKARHLYDMPD---DRVLAINHDDGVNRSIL--- 524
DB 293 FRRSDIADVE---NAIKAKLGERGYGSATVNSVDPDANKTLATLVVDAGRRLTVRQ 348
QY 525 ---GRISDAVASAVARAILPDES---ENEVIDLP---ERT---ALANRKPAD----- 564
DB 349 LRFEGNTVSAUSTLROEMKQOEGTWYNSQLVEIGKIRLDRGTGFEFTEVENRIDPINGSNDE 408
QY 565 ---VYQSKKVPRLVVEVASDKPRDGOI---GLWGSDTGTRLVTKF---EHNLI----- 608
DB 409 VDQVYKVK-----ERNTGSIINFGIGYGTEGISTYQASVKODNLTGGAASVING 457
QY 609 NRDGIQAQAEHLRSE---DKKGVKLYATKPLSLPLNDQLRATLGYYQEEVFGHS-TNGFDL 664
DB 458 TKNDGTGVNLGTEPYFTKDGVSGLGVNFENYDNSKSDTSSNKRTTYGNSVTLGFPV 517
QY 665 STRT-----LEH---EISRSIIQNGMNR---YSLRRLDLKLTQAPEPTMODLPEVDF 713
DB 518 NENNSYVGLCHTYYKIKISNFAL---YNNRLYIQSMKFKGNGIKTN-----DFDS 565
QY 714 NGKPSQEQALLAGVAVHKTVDNLVNPMRGYRORYSLEVGSSGLV---SDANMATAIRAGI 769
DB 566 FGM-----NYNSLNGYFPPTKGVKASLSGGRVITPGSDNKYYLTSADV 607
QY 770 SGVYSEFGNAGVSNRAH--QMTGCIQAGY IMSDNEHNPYRLRFPAGDOSIRGAHDSL 827
DB 608 QGFYPL-----DRDHLVAVSASAKASAGYANGFGNKRRLPFQYVTAQYIGSLRGVAYSII 660
QY 828 SP-----ISDKGYLGGOVLAGVTAE--YNEFMD-----LRLAVF 862
DB 661 CPNATVIAEYNGSGTGTFKTISSD--VIGGNALIVTASAEILLVPTPFVSDKSQNTVRTSLF 718
QY 863 GDIGNAV-----DKGFTNDTKIGAGYGVNRPASPVGVGRVNDVAVGV 902
DB 719 VDAASVNTKWKSDKNGLESDYLKRLPDYKSSKIRASTGVGFWQMSFTIGLPLVESYAKPI 778
QY 903 K--EEGNPIKLHFTITP 919
DB 779 KKYENDVEQFQSIGGSF 797

```

RESULT 13  
 US-08-942-046-2  
 ; Sequence 2, Application US/08942046

```

; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 MIS-Jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-942-046-2

```

Query Match 3.6%; Score 171.5; DB 4; Length 797;  
 Best Local Similarity 19.8%; Pred. No. 1e-05;  
 Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

```

QY 263 DVIHDLGEVYIDYRA-----VEVRGEGA-DDKAFVTVADVPLLIGDVFHHGKYEK 315
DB 72 DYKAHQEGDVLVSVAKSIISDVKIKGNSVITPEALKQNLANGKRVGDL-----IREK 127
QY 316 KLIENASAEHGFCDGRMLDRSDVY---LPDNTADVSLIYDTGQYRDEVEVFTIDPK 372
DB 128 LNEFAKSVKEHVASGVR-NATVEPIVNTLPNNRAEILIOINEDDKAKLASLTF-----K 181
QY 373 TNGLTDPDKLPVKRLEQLTLVNNGEAY-----NLQAVRALSNDLIATRYFNKAVNT 425
DB 182 GNE-SVSSSTLQEQMELQDPSMMKLMGNKFEQAQFEKDLQSTR-----DYLLNGYA 232
QY 426 EIVFPEREQIQNDQVSEFQSSSRTEPAQVD-----ESTLEPIVETVELTDG 472
DB 233 KAQITFDVQUNDEKTKVNTIDVNEGLOYDLRSARITIGNLGMASLEPLLSALHLNDT 292
QY 473 T-LMDISPIEFASNLIDQKLNVAAKARHLYDMPD---DRVLAINHDDGVNRSIL--- 524
DB 293 FRRSDIADVE---NAIKAKLGERGYGSATVNSVDPDANKTLATLVVDAGRRLTVRQ 348
QY 525 ---GRISDAVASAVARAILPDES---ENEVIDLP---ERT---ALANRKPAD----- 564
DB 349 LRFEGNTVSAUSTLROEMKQOEGTWYNSQLVEIGKIRLDRGTGFEFTEVENRIDPINGSNDE 408

```

```

OY 565 ---YOSKRPVLYYVNASDKPRDGOI---GLGMSGDTGRULYTKR---EHNLI-----608H
Db 409 VDVYKVK-----EKRSTINFGICGTGSGISTOASVKODNFLTGTAAVSIAG 457H
OY 609 NEDGYOAGAEURLSE---DKKGVKLYATKPLSHLNDLORATLIGOOEVEGHS--TNGFDL 664H
Db 458 TKNDGTSVNLGTEPEPYFTTKGVSIGGVNFEFENDNSKSDPTSSNYKRTTYTGSNVTLGFPV 517H
OY 665 STRT-----LEH---EISRSIIQONGMNRKT---YSLKRLDKLTKQAPPETWODLPVDEV 713H
Db 518 NENNSYVGLCHTYNKKTSINFAL---YNRNLYTOSMKFEKGIGITN-----DDEFS 565H
OY 714 NGKPSOEALLGAVAHKTVADNLVPMRGYRQORYSLEWSSRGVY---SPANMAIARAGI 769H
Db 566 FGW-----NYNSLRNGPYPTGCVASASLGGRTTIGSDNKKYKLSADY 607H
OY 770 SGCVSYFGDNATGNSRAH---QMTGCIQAGYIWSDNFNHVPYRLRFPFAGDOSIRGYAHDLSL 827H
Db 608 QGFPYPL-----DRDHLVVSASAKASAGYANGFENKRRLPFQYQTATGAGISLGRFAVYSI 660H
OY 828 SP-----ISDKGYLTGGCVLAVGTALE---YNPEPKD-----LRLAVF 862H
Db 661 GPNALYAEYGVGSGCTGTFPKKTISSD---YIGGAALITAAEHLVPTPEVSDKSQNTVRSLSF 718H
OY 863 GDIGNAY-----DKGFTNDTKIGAGYGVMAFSPVQOVRDVAATGV 902H
Db 719 VDASVWMTTKKSDKNGLSLDVLRKLPDYGKSSIRASTGVGQWQMSPIOLPLFVSTAKPI 778H
OY 903 K--EGGNPKIHLFFICTPFE 919
Db 779 KKEYDVEBOFOFISIGSF 797

```

RESULT 14  
 US-08-942-046-4  
 Sequence 4, Application US/08942046  
 Patent No. 6264954  
 GENERAL INFORMATION:  
 APPLICANT: CHONG, Pele  
 APPLICANT: THOMAS, Wayne  
 APPLICANT: YANG, Yan Ping  
 APPLICANT: LOOSMORE, Sheena  
 APPLICANT: SIA, Dwo Yuan Charles  
 APPLICANT: KLEIN, Michel  
 TITLE OF INVENTION: HAEMORRHOUS OUTER MEMBRANE PROTEIN  
 NUMBER OF SEQUENCES: 55  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sim & Mcburney  
 STREET: 6TH Floor, 330 University Avenue  
 CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R7  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/942,046  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/433,522  
 FILING DATE: 12-SEP-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: STEWART, Michael I  
 REGISTRATION NUMBER: 24,973  
 REFERENCE/DOCKET NUMBER: 1088-732 MIS:jdb  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (416) 595-1155  
 TELEFAX: (416) 595-1163

```

; INFORMATION FOR SEQ ID NO: 4:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH: 797 amino acids
;         TYPE: amino acid
;         TOPOLOGY: linear
;     MOLECULE TYPE: protein
; US-08-942-046-4

```

Query Match	Score	DB	Length
3.68;	171.5;	4;	797;

Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

QY	263	DVITHDGEPEYIDYRA-----VEVGEGA-DDKAPTTVADEVPLIGVFNHGKYEK	315
Dp	72	DYKAHQESDVLVSVAKSIISDVKIKGNSVITPEAKKONLDANGKPVGDV----TREK	127
QY	316	KNLIENSAEHGYDGRHLRSVVI---LPDONTADVSLYDGTQYRFDEVEFTIDPK	372
Dp	128	LNFEAKSVKEHYASVGR-NATVEPIYNTLPNNAEILLIQINEDDKAKLSLTF-----K	161
QY	373	TNOLTTDPDKLPVKRELLBOLLVYNNGBAY-----NLOAVRLSNDLIATRYENYNT	423
Dp	182	GNE-SVSSSTLOEBOMELQDPSMWKMLCKNKEGAGOFERKLQISIR-----DYLLNNGYA	232
QY	426	EIVPERQIONDPOVSFEQSSSKTEPEAOVD-----ESTLPEVIEVETLDC	472
Dp	233	KAQTKTDVQULNDKTKYVNVYIDVNEGLQTDLSARITICNLGMSALEBPLLSALHNDT	292
QY	473	I-LMDISPIEFSASNLIODKILNVAAKARHLYPD-DRVLAINHDDGVNRSIL---	524
Dp	293	FRSDIDAVE---NAIKAKLGEYGSAFVYNSVPDDEANKTLATILVVDAGRLVRO	348
QY	525	---GRISDAVSAVARILDEDES---ENEYIDLPE---ERT---ALANKTPAD-----	564
Dp	349	LREGEYTVASDSTLRQEMROQEGTWNYSOLVELKIRLDRTGEFEYENKIDRINGSNDE	408
QY	565	---VYQSKKPYLVFVASDKPRDQI---GLGMSDGTSTRLVTRF-EHNL-----	608
Dp	409	VDVYVYKK-----ERNGTSGINFGITGTEESISTQASVKODNPLGTGAANSIAG	457
QY	609	NRDGYOAGELRLSE---DKGYKLYATKPLSHPLNDOLRATLGYOQOEVGHS-TNGFDL	664
Dp	458	TKNDYGTSVMLGYTEPEFTKDGVSIGGNVFEFENDNKSQTSNRYKRTYGSNVTLGPFV	517
QY	665	STR-----LEH---ELSRSTIQNGGNRT---YSLRYRLDKLAKTOAPETWODLPPYDV	713
Dp	518	NENNSYVVGHTHYNKISNFLE---YNRRLYIQSMKFKNGIKTN-----DEFES	565
QY	714	NGKPSQFALLAGVAHVHTVADNLVNPWRGYRQRYSLSEVGSSGLV---SPANNAIRAGI	769
Dp	566	FGW-----NNLSNLKRGFPFKKGYKAKASLGAGYVTLPGSDNNKYYKLSADY	607
QY	770	SGYVSFEEDNAYGSNRAH---OMTGGIOAGYIWSDNFNHVPYRLRFEPAGQDOSIRGVAHSL	827
Dp	608	QGEFPL-----DROHLWYVASAKASAGYANGCNKRRLPEYQYTTAGIGISLGFAYGSI	660
QY	828	SP-----ISDKGYLVLTGOVLAVGTAE---VYEEPMKD-----LALAF	862
Dp	661	GPNALTYAETGNGSGTGEFKKTISSD-VIGGNALATATASALIEVTPPVSDKSONTVFTSLF	718
QY	863	GDIGNAY-----DKGFTDTRIGAGVGRNAPSVPQGVVVDVATV	902
Dp	719	VDASAVWNTKWKSDKNGLSDVLAKRLPDYKSSRIASVYGVQWQSPICGPLVFSIAKFI	778
QY	903	K--BEGNPJLHFFIGTGP	919
Dp	779	KXYENDVEQFQSIGSP	797

RESULT 15  
US-08-942-046-6  
; Sequence 6, Application US/08942046  
; Patent No. 6264954



```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 393-2000
TELEFAX: (415) 393-2286
TELEX: 340817 MACPAG SFO
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3031 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-689-008-2

Query Match      3.4%; Score 159.5; DB 1; Length 3031;
Best Local Similarity 19.2%; Pred. No. 0.001;
Matches 201; Conservative 131; Mismatches 360; Indels 357; Gaps 51;

QY 43 PAHDPAI-----NQAKAGNPVLLT-PRQIQARL--NAAGLNAPQSOAL- 84
DB 1720 PPHSLAVEYQTMAGVPAQMDQARAGIAGVNASNPQDYRQALAFQAALTYNTSTRMGLT 1779
QY 85 ---DVVNFDDQSPI-SRIGESPPGLID-MSVIEETTPLEELFAQESTEMGINNDYI 139
DB 1780 RLKDIQSFSSQAPVEAAAAQSTRLSLMVPVETQPLMEQWLSA-----HPND-- 1829
QY 140 PEYQGEQPNSEVVVPTLEPEKPKGLIKRLYARLFNDGVNKKVPRLKAKFYQSSQSGETSAT 199
DB 1830 -----TALREHMLHPGCGPPDKAGLARQ-----AGYQQLNAGRLAA 1866
QY 200 GSSHQKTEPYANIKALEITQESANDLNGSIPRLKOTALVAAAVGVYDIDISTIRNST 259
DB 1867 EOSFQ-----SALQINS-----HDAD-----SL 1884
QY 260 GEVDYIIHDLGEPYIIDYRAVEYREGADDKAFTVADEVPLLIGDFHNGKETKKNL 319
DB 1885 GGMGLVSNMGQDTA-EARRYEEEAADPK-----TADKMRPALAGMAVSGEYASVROL 1938
QY 320 ENASAEHGTFDGRW-----LDRSVDYI----- 341
DB 1939 ---AAHOYTEAKQOLATLARQGYTGATLMLADLQSTGQIAAAEQRYGLISREPNN 1994
QY 342 -----LPNDTADV-SLIYDTGYQ--RDEYVFFTIIDPKTNOLTTDPDKLPVK 386
DB 1995 OLALMGLARVDMAQNGTAEARQLLSRVGPQYASQVGEIEVSGIMAAASQTSARAKVSI 2054
QY 387 RELLEQL-----LVNMGAEVYNIQ-----AVRAL-----SND 413
DB 2055 REAMQAARDPVVRINTLANALQOQSDVAEAGRVMOPIIANPYTAODRQALITYTGGSGND 2114
QY 414 LIATRYFNKVNTEIYFPEREQIONDOVSFEQSSSRTEPAQVDESTLEPVIETVELTDCI 473
DB 2115 AMTRQLLAGSPADYSPAIRSIA-EEMEIKODLARLSVNSPVPILIRALQOPDPTGAR 2173
QY 474 LMDISPIEFSASNLQDKLNLVAAKARHLYDMPDRV-----LAINHDDGVNRSILGRI 527
DB 2174 GVAAVADFERQGDVYHAMMARLIASTRITIDLSPDORLSYATEYMKISNPVAAAR-LIAPL 2232
QY 528 SDAV-SAVARAIIIPDESE-----NEVIDLPERTALANKRTPADVQSKVPLVYVVASPK 581
DB 2233 GDSGSAAGSALLPEQVYTTLOQLRMGISVYASQDLNQRGDOQAQVADHLAPALQADPEAT 2292
QY 582 PR-----DQIGLGWSDTGTRLYTKFEHNLINRDYQAGAEALRSEDKKV--KL 630
DB 2293 KFLALRLYNGHGRKPKALEID-----LAVLRHNQDDADQAQAAVQAAVNSHNSLARL 2347
QY 631 YATRPLSLPLNDOLRATLGYQOEYFGHSTNGFDLSTRTLEHISHSITONGGWNPTYSRL 690
DB 2348 AMDGVQSSPMDARMLAAVADQADGH-----GQRTIE-----DLRRAYDLR 2389
QY 691 YRL-----DKLKTQAPET-----MODLPVDVYNGKPSQA----- 721
DB 2390 LQOYEGTRRAASGAAQOEDAL--APSPINPFRPRRGYGHOTELGAPVYTGGSYSABAASPD 2446
QY 722 ---LLAGVAVH-KTVADNL---VNPNGRYQRYSLSEVGSSGLIVSDANMALA-----RAG 768
```

```
DB 2447 TSDQMLSSIAGQIRTLLENLAPSIDGLGFRSR-SGEHG-MGRLEENAPIVGLRPLQAG 2504
QY 769 ISGYVSFGDNNAVGSNNRAHQMTGGIQAGYIMSDNFN-----HVPYRLFFAGDQDSIRGY- 822
DB 2505 ASA-----LFSITPTMWSGNLNTSGSYVDV--RY--GTMMGVQAYN 2543
QY 823 AHDSLSPIQSKYLTGG---QVLAVGTAEVNEYEFMKRLRLAEFGDIGNADKGTNDTKI 879
DB 2544 QYDS-----TYNMGDDQORIAAGTAE-----AGFAPDVQF 2573
QY 880 G-----AGVGRWASPVPYGOVRVDVATGVC 903
DB 2574 GNSWVRADVG---ASPIGFPITNVLGVE 2599

RESULT 17
US-09-268-347-49
; Sequence 49, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 2314
; TYPE: PRP
; ORGANISM: Hemophilus influenzae
US-09-268-347-49
```

```
Query Match      2.9%; Score 138; DB 4; Length 2314;
Best Local Similarity 19.3%; Pred. No. 0.042;
Matches 189; Conservative 132; Mismatches 315; Indels 342; Gaps 53;

QY 68 ARLNAAGLNAPQSOA-----LDVNFDDQSPISRIGESPPGLDMSVIEETTPLS 119
DB 435 AELQSGGLFTFTPTNASTKTYVGTDLKFTDMS-----NTALEDTTTRIT 479
QY 120 LEEEL-PAQESTEMGINPNDYIPEYQGEQPNSEVVVPTLEPEKPKGLIKRLYARLFNDGVN 178
DB 480 KDKIGFSNKA-----GTYDENKPYLDK----- 501
QY 179 KYPRLEKAKFYQSSQGET--SAIGSSHQKTEPYAN-IKALEITQESAMDLNGSIPRLR 235
DB 502 --DKLKVGSSTLNGGLTYNNNTIGSNNKOIOVGADGDKFA-----DVNVNVSMAA 549
QY 236 Q--TALVAAAVGYDIDISIRNSIGEVDYIIHDLGEPYIIDYRAVEY-----RGE 285
DB 550 KFGTTRITREEIGFADAD-----GKYD-----KKSPIYDKKQLOVYGCVKITYKDSGI 595
QY 286 GADDKAFTVADEVPLLIGDYVFNHKGYTEFKKMLIENASAEHGTFDGRWLDRSDVYLLPN 345
DB 596 NAGDQKISNVKQAT-----DDTDAVYKQIKOVQDA-----DALQSFST--RDE 639
QY 346 TAD---VSLIYDTG--TQYRFDEVF-----FTIDP--KTNQLTTD 379
DB 640 KQOETISMLYNSNGTPTNFTTTFRAGENGISISNDIAKGVKVGIDPITNGLTTPRLTVG 699
QY 360 PDKLPVKRELEQLLVN-----MGEAVNLQAVR-----AL 410
DB 700 SPDKDQTOLVIEQVAVSNGTKNIIINGVSPTLPSIINAGSVRTTEGNTTTSDEDKSKAAS 759
QY 411 SNDLIATRYFNKVNTEIYFPEREQIONDOVSFEQS-----SSRTEPAQVDESTLEP 462
DB 760 IGDILINTG-FNLKN-----NSNSYGFVSTYNTVDFIDGNATTAKYVYDETN-- 804
QY 463 VLEVEYELFDGIIMDISPIEFSASNLQDKLNLVAAKARHL-----YDMDDDRVL 511
DB 805 --QTSKVITYDVAVDEKTIETLIGDN--GTYNKIGVKTITTTTNTANNGATYNSTTDNDAL 859
```





Db 1017 SNTLANTNDKG-----SVRTTEOG-NIIKDEKTRAISIVDLSAGFNLOGNGEAV 1067  
 QY 870 DKGFTNDT 877  
 Db 1068 DFVSTYDT 1075

# RESULT 19 US-08-945-567D-4

; Sequence 4, Application US/08945567D  
 ; Patent No. 6448386  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SASAKI, Ken  
 ; APPLICANT: HARKNESS, Robin E.  
 ; APPLICANT: LOOSMORE, Sheena M.  
 ; APPLICANT: CHONG, Pele  
 ; APPLICANT: KLEIN, Michel H.  
 ; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF  
 ; TITLE OF INVENTION: MORAXELLA  
 ; FILE REFERENCE: 1038-745 MIS  
 ; CURRENT APPLICATION NUMBER: US/08/945,567D  
 ; PRIOR FILING DATE: 1996-04-29  
 ; PRIOR APPLICATION NUMBER: 08/431,718  
 ; PRIOR FILING DATE: 1995-05-01  
 ; PRIOR APPLICATION NUMBER: 08/478,370  
 ; PRIOR FILING DATE: 1995-06-07  
 ; PRIOR APPLICATION NUMBER: 08/621,944  
 ; PRIOR FILING DATE: 1996-03-26  
 ; PRIOR APPLICATION NUMBER: PCT/CA96/00264  
 ; PRIOR FILING DATE: 1996-04-29  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 4  
 ; LENGTH: 1833  
 ; TYPE: PRT  
 ; ORGANISM: Moraxella catarrhalis  
 ; US-08-945-567D-4

Query Match 2.9%; Score 136; DB 4; Length 1833;  
 Best Local Similarity 19.7%; Pred. No. 0.042; Indels 298; Gaps 49;

Matches 191; Conservative 128; Mismatches 351; Indels 298; Gaps 49;  
 QY 22 PLMTSQAALQOONNPAIIINHVAHDPAI--NOAKAGNPVLLTFPEQIOA---RLNA-AGLN 76  
 Db 294 PYLDKQKLVKGVSAITIDGIDAGNKKISLAKGSSANDVAITEQLKAAPTLNAGGIS 353  
 QY 77 AKFQSAALVAVNDDQSPISRIGESPPGLD-----MSVIEETPLST--EELFAQEST 129  
 Db 354 VPTETISVDKSGNVAPTYNIGVKTTELNSDGTSDKFSVKGSGTNNSLVTAHSLAYLN 413  
 QY 130 ENGINPND-----YIPEYGEOPNSEVVPPTLEPEKPGIKRLKRLARLPDGVNKKPRK 184  
 Db 414 EVNRRTADSALOSTVVEEDDDDAATVAKDTRY-----NAGAVSILIKLK 458  
 QY 185 AK-----FYOSQSGETSAIGSSHQKTEPYANIKALEDLTQESA-----M 225  
 Db 459 GKNGLTAVATKKGDTVFFGLSDPSGLT--IGKSTLNDG--LTVAKDTEQI--QVANGIKFT 514  
 QY 226 DLNGSLP--RLKQTAALVAAVAGYVIDLSITRNSTIGEVDVYIHLGEPYIDYRAEV- 282  
 Db 515 NVNGSPGIGTANTARITRDKIGFAGSDGAV-----DINKP--YLDQDKLQV 560  
 QY 283 -----RGGADKAFATVADVEPLL-----IGVFHHG-K 311  
 Db 561 NVKITNTGINAGKALTGSPPLPSIADOSSRNIELGNTIQDKKSNASINDILMTGFN 620  
 QY 312 YETKKNLLENASAEHGYFDGRMLDRSDVY--LPDNTADVSLITDTGTQYRDEVEVF-ET 368  
 Db 621 LKNNNPIDFVS-----TYLDVDFRANGAATTAITYHTDAN--KTSKVYVDVA 665  
 QY 369 IDPKTNQL--TTDDPKLPVKRELLLEQLLTVNNGEAVYNLQAVRALNDLITATRY-FMNVNT 425

Db 666 VDDTTIHLTGTDNKKLGKVTTKLN-----KTSANGNTATNFVNSDE 709  
 QY 426 EIVPEREQION-DQVSFEOSSSSTPERAOYDESTLEPIETVELTDCIIMDISIEESA 484  
 Db 710 DALVNAKDIAENLNTLAKETHTTKGTADPALQTFVKKVDENNADNATTVG--QKNA 767  
 QY 485 SNLIQDKLNLVAAKARHLHYDMPDR-----VLAINHDDGVNSILGRISDAVSAVARILP 540  
 Db 768 NN-----QVNTLTILGGENLNKTDKNGVTGINTSGLK-----AGKSTLN 810  
 QY 541 DESENEVIDPERPALANRKTTPADYQSKVPLVYFVASDKPRDQIGLGNGSDPTGRLV 600  
 Db 811 DGG-----LSIKNPTGSEQIQVADGVKPAV-----NNNGVYAG--IDGTTRI 853  
 QY 601 TKFPHNLINRCGYAGAEALRISEDKKVKLYATKPLSHPLDQLRAITGYQOEVPFGH-ST 659  
 Db 854 TRDEIGFTGNG-----SLDK-----SKP-----HLK 876  
 QY 660 NGFDLSTRTLEHEISRSTION-----GNNRTYSLRYRLDKLKTQAPPTWQDLPVDFV 713  
 Db 877 DGINAGKKKITINIOSGETAQNASHAVYGG--KIYDLKTELE----- 915  
 QY 714 NGKPSQALLAGVAVHK-TVAD-----NLVNPARGYRQARYSLEY---GSSGLVSDAM 762  
 Db 916 -NKISSPAKTAQNSLHERSVADDEQNNFTVNSPYSDTSTQSVITPAGENGITTKYK 974  
 QY 763 AIAAGISGYVSFEDNAGSRAHOMTGIGIAGIYSDNFHVPRLRFFAGQDOSIRCY 822  
 Db 975 GVAVRGI-----DQTKGLTTPRLTYGNNGKVIDISON-----GNTITIGL 1016  
 QY 823 AHDSLSPISDKGYLTGGOVLAFTAENYEMKD---LRLAVFGDI-----GNAY 869  
 Db 1017 SNTLANTNDKG-----SVRTTEOG-NIIKDEKTRAISIVDLSAGFNLOGNGEAV 1067  
 QY 870 DKGFTNDT 877  
 Db 1068 DFVSTYDT 1075

## RESULT 20

US-08-621-944A-3  
 ; Sequence 3, Application US/08621944A  
 ; Patent No. 6440425  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SASAKI, Ken  
 ; APPLICANT: HARKNESS, Robin E.  
 ; APPLICANT: LOOSMORE, Sheena M.  
 ; APPLICANT: KLEIN, Michel H.  
 ; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER  
 ; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sim & McBurney  
 ; STREET: Suite 701, 330 University Avenue  
 ; CITY: Toronto  
 ; STATE: Ontario  
 ; COUNTRY: Canada  
 ; ZIP: M5G 1R7  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/621,944A  
 ; FILING DATE: 26-MAR-1996  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/478,370  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Stewart, Michael I

```

1 REGISTRATION NUMBER: 24, 973
2 REFERENCE/DOCKET NUMBER: 1036-587
3 TELECOMMUNICATION INFORMATION:
4 TELEPHONE: (416) 595-1155
5 TELEFAX: (416) 595-1163
6 INFORMATION FOR SEQ ID NO: 3:
7 SEQUENCE CHARACTERISTICS:
8 LENGTH: 1992 amino acids
9 TYPE: amino acid
10 STRANDEDNESS: single
11 TOPOLOGY: linear
12
13 OS-08-621-944A-3

```

Query Match	2.98; Score 136; DB 4; Length 1992;
-------------	-------------------------------------

QY	22	PLMTSOLAQONNPNANIINHVAHDTAI-NQAKAGNPVLLTPEIDIA----	RLNM-AGLN	76
Db	453	PLYLKKQLKVGVAITIDGIDAGKKKISLNKKSSSANDAVITIEQKAKAPLLNAGAGIS		512
QY	77	AKPOCALDVNFFDOOSPISRIEOSPPLGLD-----MSVIEETPLSL-EELFAOEST		129
Db	513	VTPTEISVDASGNAVTAFTYNGVATTELSNDSGTSODKSVKSGSNNLSLVTAHEHLASYLN		572
QY	130	EMGINPND-----YIPEYTOGEOPNSEVYVPPPLLEPEKPOLIKRLYARLFENDGVNKKVPRLK		184
Db	573	EYNRTADASALOSFYKAEEDDDANAITVAKDSTK-----NAAVASILTKLK		617
QY	185	AK-----FQSSOSGEMSAIGSSHOKTEPEYANIKALLEDITQESA-----M		225
Db	618	CKNGLTVATKKDGYTFYGLSDOSGLT--IGKSTLNNDG-LIYKDTNEQI-QVGGANGIKFT		673
QY	226	DUNGSI-P-RLKOTALVARAVGYVIDLSIIRNSIGEVDIITHDLEGPVYIDYRAVEY-		282
Db	674	NVNGSNPFGTIANTRITRDKIGFAGSDGAV-----DTNKP-YLDOOKLOY		719
QY	283	-----REGADDAKFTYVADEVPLL-----IDGVFHHG-K		311
Db	720	NVKTNTNTINMGKAKITGSLPTLPIADOSSNIELGNTIODKDSMAASINDLITATGNN		779
QY	312	YETKKNLIEENASAEHGFYDGRWLDRSDVYI--LPDNTADVSLIYDTGTQYREDEVYF-FT		368
Db	780	LKNNNPIDFYS-----TYDLYDRANGCNATATYTHDTAN--KTSKVYVDYN		824
QY	369	IDPKTNOL--TTDPKLPYKRELEBOLLTVNMGAYNLQAVALSNDLIATRY-FNMVNT		425
Db	835	VDDTTHILGTDDNKKLGKTTKLN-----KTSANGNATATFNNSSDP		868
QY	426	EIVFEREDION-DOVSFEQSSSRTEPAQVDESTLEPIETVELLDGILMISPISFEFA		484
Db	869	DALVAKKDAEULNTLAKEIHTTKGTADTALOTFTYVKKYDENNMADANAITVG--QKKA		926
QY	485	SNLIDOKDLNVAAKARHLHYDMPDR-----VLAINHDDGVNRSILGRISDAVASAVARAII		540
Db	927	NN-----QVNTLILKGENGLINKTDKNGVTFGINTTSGLK-----ACKSTLN		969
QY	541	DESENEVIDLEPRTALANRKTTPADYVQSKKVPYLVFVASDKRDRDQIGIGWGSDTGTRIV		600
Db	970	DGG-----LSIKNPTSEQIQVGADGVCKFAKV-----NNNGVYVAG--IDGTRRI		1012
QY	601	TKFEHNLLRPDQYQAGAELELSEDKKGVKLYATKRPLSHPLNQDLBATLIGYQGEVFGH-ST		659
Db	1013	TRDELGFPGTNC-----SLDK-----SKP-----HLSK		1035
QY	660	NGFDSLSTLLEHISRSITQ-----GGMNTRYSLRYRLDKLKTQAPPETWQDLPVDEV		713
Db	1036	DGINAGCKKITINIGCEIAONSHDAVTGG--KIYDKTELE-----		1074
QY	714	NGKPSQELLALGAVAVHK-TVAD-----NLVPMNGYRQRYSLVY-----GSSGLVSDAMN		762
Db	1075	NKLSSTYKTKTQNSLHESVADQENNFYVSNPYSSTYDTSKTSVDITFAGENGIGITTKVKR		1133

OY	763	AIAAPAGISGVSPSGDNNAYSNRAHQMTGTGLQAGVIMSDPNEHNPVRILAFPGGGOSIFGY	822
Dd	1134	GVAHVGI-----DQTGLTTPKLVGANNNGKGIYDSQN-----GQNTITGL	1175
OY	823	AHMSLSPISKDYITLGGOVLAVCTAETNYEPMKD--LRLAVFGDI-----GNAY	869
Dd	1176	SNLLANTLTNRKG-----SVRTTEQG-NIINKDEKTRASTIVDVLSAGFNLGGNGEAV	1226
OY	870	DKGEPTNDT	877
Dd	1227	DFVSTYDT	1234

RESULT 21  
US-08-945-567D-3  
; Sequence 3, Application US/08945567D

Query Match	2.9%	Score 136;	DB 4;	Length 1992;
Best Local Similarity	19.7%	Pred. No. 0.049;		
Matches 191;	Conservative 128;	Mismatches 351;	Indels 298;	Gaps 49;
QY	22 PLMTSOLAQQONNPANINHHVPAHDTA--NQAKACNPVLLPPEIOIA---RLNA--AGLN 76	: : :	: : :	: : :
Db	453 PFLDKQLKVGSAVATTINGIDAGKKKISLNLAGSSANDAVITEQLKAAPLLNMGAGIS 512	: : :	: : :	: : :
QY	77 AKPOSQALDVNFFDDOSPISRIGEOSPPLGLD-----MSVIEETTPPLST--ELFPAQEST 129	: : :	: : :	: : :
Db	513 VPTPEISYDAKSGNVATAPYINIGVKTTELSNGTSDKFSVKSGSGTNNSLVTAEHLASYLN 572	: : :	: : :	: : :
QY	130 EMGINPND-----YIPEYQGEOPNSEVYVPPVLEEKPEQLIKRLXARLENDGVNKVPRLK 184	: : :	: : :	: : :
Db	573 EVNRTADASLOSFTYKEEDDDANAITAKDTK-----NAAVSAILUK 617	: : :	: : :	: : :
QY	185 AK-----FYOSSQSGETSAIGSSHOKTEPAPANIKALEDTIOESA-----M 225	: : :	: : :	: : :
Db	618 GKNGLTVAATKKGIVTFYFGLSQDSGLT--IGKTTLNNDG--LYAKKDINEQI--QVGGANGIKFT 673	: : :	: : :	: : :
QY	226 DINGSIP--RLRQALVAVARAVGYVDIDISITIRNSIGEVDYIIHDLGEBEYVYIDYRAVEV- 282	: : :	: : :	: : :
Db	674 NVNGSNPQGIANTFARITRDKIGFAGSGDAV-----DTNKP--YLDQOKLQVW 719	: : :	: : :	: : :
QY	283 -----REGAGDDKAFYTADEVEPLL-----IGDVEHHG-K 311	: : :	: : :	: : :
Db	720 NVKITTNTGIMAGKCAITGTSPLTPIADOSSKNIELGNTIQDKDKSMAASINDILLTGCEN 779	: : :	: : :	: : :
QY	312 YETKKNLIENASAEHGVEFGRLDRSSVDY--LPDNTADVLSLYDTGTQYRDEVEVF--FT 368	: : :	: : :	: : :

```
Db      780 LKNNNNIDFVS-----TYDIDVFANGNATTAATVTHDTAN--KTSKVVYDVN 824
Qy      369 IDPKTNOL--TTDPDKLPYKRELEQLLTVMGSAVNLQAVRALSNDLIATRY-FNNVNT 425
Db      825 VDDTTIHLTGTDNRKRLGVTKTKLN-----KTSANGTATNFVNSSDE 868
Qy      426 EIVPEREQLON-DQVFEQSSSRTEPAQVDESTLEPVIEVELTGILMDISPIESA 484
Db      869 DALVNAKDIAENLNTLAKELHTTKGTADTAALQTFYKKBVDENNADANATTVG-QKNA 926
Qy      485 SNLIQDKLNLVAAKARLLYMPDR-----VLAIHDDGVNRSILGRISDAVASAVALP 540
Db      927 NN-----QVNTLLTKGEGNLKIKTDKNGTVTFGINTSGLK-----AGKSTLN 969
Qy      541 DESENEVIDPERTALANRKTADVYOSKRVPLVFVASDKPRDGOIGLWGSPTGRV 600
Db      970 DGG-----LSIKNPTGSGQIOVGADGVKFAV-----NNGGVGAG---IDGTTRI 1012
Qy      601 TKFEHNLINRDGYQAGAELELSEDKKGVKLATKPLSHPLNDLRAVLGYQOEVEFGH-ST 659
Db      1013 TRDEIGFTGNG-----SLDK-----SKP-----HLSK 1035
Qy      660 NGFDLSTRTEHEHSKSIQN-----GGMNRTYSLRYRLDKLKTQAPPETWODLPVDFV 713
Db      1036 DGINAGKKTITNIOGSEIAQNSHDAVTG--KIYDLKTELE----- 1074
Qy      714 NGKPSQALLAGVAVHK-TVAD-----NLVPMRGYRORYSLEV-----GSSGLVSDAM 762
Db      1075 -NKISSTAKTAKNSLHEFSVADEQGNFTVSNPYSYDTSKTSVITFAGENGITTVYVK 1133
Qy      763 AIAKAGISGVYSEGDNAVGSNRAHQMTGIIQAGYIWSDNFNHVPYRLRFAGDOSIRGY 822
Db      1134 GAVRVGI-----DQTKGLTTPKLTGVNNGKGIVIDSQN-----GQNTITGL 1175
Qy      823 AHSLSPLSKGVLGTGOVLAVGTAENYEFMKD--LRLAVFGDI-----GNAY 869
Db      1176 SNTLAVNTNRK-----SVRTTEGG-NIKEDKTRAASIVDVLSAGFNLOGNEAV 1226
Qy      870 DKGFINDT 877
Db      1227 DFTVSTYDT 1234

RESULT 22
US-09-268-347-48
; Sequence 48, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Locomore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 2048
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-48
```

Query Match 2.9%; Score 136; DB 4; Length 2048;

Best Local Similarity 19.6%; Pred. No. 0 051;

Matches 188; Conservative 133; Mismatches 356; Indels 284; Gaps 48;

```
Qy      22 PLMTSGALAOQNPNANIINHVPADTAI-NOAKAGNPVLLTPEQIA--RLNA-AGLN 76
Db      508 PYLDKQLVGVSAITIDGIDAGNKKISMLAKGSSANDAVTEQLKAAPLTINAGAGIS 567
Qy      77 AKPQSGALDVVNDDQSPISRTGQSPPLGLD-----MSYIEETPLSL--EELFAQEST 129
Db      568 VTPTEISVDKSGNVTAPYTNIGVKTTELSNDSKFSVKGSGTNNSLVTAHLASYN 627
```

```
Qy      130 EMGINPD-----YIPEXQEQPNSEVVPPLEPEKGLIKRLYLRLFNDGVKKVPLK 184
Db      628 EVNRTASALQSFYKKEDEDDANAIYAKPTK-----NAGAVSIILKLK 672
Qy      185 AK-----FYQSSQSETSAGSSHOKTEPYANIKALDEDTQESA-----M 225
Db      673 GKNGLVATKKGDTVFTEGLSDSGLT--IGKSTLNNDDG-LVFKDPTNEQI-QVANGIKFT 728
Qy      226 DLNCSIP--RLRQALVAARAVGYIDDLSIIRNSIGEVYIINDLEPVYIDVRAEV- 282
Db      729 NVNGSNPQGTGIANTARTTRDKIGFAGSDGAV-----DTNKP-VLDDDKLQVG 774
Qy      283 -----RGEADDAKFTTVADEVPL-----IGDFVHHG-K 311
Db      775 NVKTTNGINAGGAATIGLSPTLSIDQSSRNIELNTIODKKSNAASINDLINLNGFN 834
Qy      312 YETKKNLIENASAEHGYFDGRMLDRSYDT--LPDNTADVLSLYDTGTQYRFDEVF-FT 368
Db      835 LKNNNPDIFVS-----TYDIDVFANGNATTAATVTHDTAN--KTSKVVYDVN 879
Qy      369 IDPKTNOL--TTDPDKLPYKRELEQLLTVMGSAVNLQAVRALSNDLIATRY-FNNVNT 425
Db      880 VDDTTIHLTGTDNRKRLGVTKTKLN-----KTSANGTATNFVNSSDE 923
Qy      426 EIVPEREQLON-DQVFEQSSSRTEPAQVDESTLEPVIEVELTGILMDISPIESA 484
Db      924 DALVNAKDIAENLNTLAKELHTTKGTADTAALQTFYKKBVDENNADANATTVG-QKNA 981
Qy      485 SNLIQDKLNLVAAKARLLYMPDR-----VLAIHDDGVNRSILGRISDAVASAVALP 540
Db      982 NN-----QVNTLLTKGEGNLKIKTDKNGTVTFGINTSGLK-----AGKSTLN 1024
Qy      541 DESENEVIDPERTALANRKTADVYOSKRVPLVFVASDKPRDGOIGLWGSPTGRV 600
Db      1025 DGG-----LSIKNPTGSGQIOVGADGVKFAV-----NNGGVGAG---IDGTTRI 1067
Qy      601 TKFEHNLINRDGYQAGAELELSEDKKGVKLATKPLSHPLNDLRAVLGYQOEVEFGHSTN 660
Db      1068 TRDEIGFTGNGSLDKSRPHLSKD--GINA-GGKKTINIOSGELQA-----N 1111
Qy      661 GFDSLSTRTEHEHSRSITONGGMNRTYSLRYRLDKLKTQAPPETWODLPVDFVNGKPSOE 720
Db      1112 SHDAVT-----GG--KIYDLKTELE-----NKISST 1135
Qy      721 ALLAGVAVHK-TVAD-----NLVPMRGYRORYSLEV-----GSSGLVSDANMAIARAGI 769
Db      1136 AKTAQNSLHEFSVADEQGNFTVSNPYSYDTSKTSVITFAGENGITTVYKNGVVRVGI 1195
Qy      770 SGVYSEGDNAVGSNRAHQMTGIIQAGYIWSDNFNHVPYRLRFAGDOSIRGYAHDLSLP 829
Db      1196 -----DQTKGLTTPKLTGVNNGKGIVIDSQN-----GQNTITGSLNTLAV 1237
Qy      830 ISDKGYLTGGOVLAVGTAENYEFMKD--LRLAVFGDI-----GNAYDKGFINDT 876
Db      1238 TNDKG-----SVRTTEGG-NIKEDKTRAASIVDVLSAGFNLOGNEAVDFVSTYD 1288
Qy      877 T 877
Db      1289 T 1289
```

RESULT 23

US-09-134-001C-3159

; Sequence 3159, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

;; PRIOR APPLICATION NUMBER: US 60/064,964  
;; PRIOR FILING DATE: 1997-11-08  
;; PRIOR APPLICATION NUMBER: US 60/055,779  
;; PRIOR FILING DATE: 1997-08-14  
;; NUMBER OF SEQ ID NOS: 5674  
;; SEQ ID NO 3159  
;; LENGTH: 10182  
;; TYPE: PRT  
;; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3159

Query Match 2.9%; Score 135; DB 4; Length 10182;  
Best Local Similarity 20.0%; Pred. No. 0.93;  
Matches 166; Conservative 106; Mismatches 293; Indels 266; Gaps 38;

QY 27 CALAQQNNPANI-TINHPADHTAINQAKAGNPVLLTPPEIQARLNAAGLNAKPOSQLD 85  
DB 2695 QAAEQAIQHANSVINNGDATSQINDAK--NTVEQAQRDYEAQSMLRADKSQLOS-AVD 2751  
QY 86 VVNP-----DOSP-----ISRIEESPPLGLDMSYEETPLSLEELAEQESTENG 132  
DB 2732 TLNRDVLTKDKKRPASVRRYNEAISIRKELDTAKADASTLRNTPSQVBDALNKINT 2811  
QY 133 INP--NDYIPEYQEQPNSEVVPPLEBEKPLIKRLVRLFNDGVNKPRLKA----- 185  
DB 2812 VQPKVNAIALLQPKENNELVQAK-----KRL-----QDAVNDIPQTQMTQOT 2856  
QY 186 -KFTYSSQSGTSAIGSSHQKTEPYANIKALEDITQESAMDLNGSIPRLROT--ALVA 242  
DB 2857 INNYNDKQREAFERALTSAQRVID--NGDATQETISEKS-----KVEQAMQALNTA 2905  
QY 243 RAVGYDDIDLSIRNSIGEVYIHDLEPVYIDYRAVEVRGEGADKAFITVADEVPLL 302  
DB 2906 K-----SLUR-----YETKKNLEINASAHEGYFDGRWLD-----ADKNELOTAYNK---L 2923  
QY 303 IGDVFNHCK-----YETKKNLEINASAHEGYFDGRWLD-----RSYDVIL 342  
DB 2924 IENVSTNGKKPASIRQYETAKARIQNOINDAKNEARILGNDNPQVSOVTOALNKIKAIQ 2983  
QY 343 PDNTADVLSLID-----TGQYRFDEVYFTTIDPKTNQLTDD-----PDKLVKREL 389  
DB 2984 PKLTEAIIIMLQNKENTELVNAKNRLENNAVNDT--DP--THGMTQETINNYNAKKREAQNET 3041  
QY 390 LEOLLTVNMGCA-----YNLOAVRALSNDL-----LATREFMN---VNFET 427  
DB 3042 QKANNIINGDATADISSEKSYEQVLOALQNAKNDLRADKRELQTAIVNKLITQNVNTNG 3101  
QY 428 VFPE-----REQIQNDQVSEFQSSSSRTPEPAQVDESTLEPIYETVELFDGILMDISP 479  
DB 3102 KKPSSIQNYKSAARIEN-----QYNTAKNEAHNVLENT--NPVTNAVEDALRKINAIO 3154  
QY 480 IEFASNLIOCK-----LNVAAKARHLVMPDDR 509  
DB 3155 EYTAIKNILQCKDENSELVRAKEKLQDAINSOPSLNGMTOESINNYTTKRRRAQNIASSA 3214  
QY 510 VLAINHDDGVNRSTLG---RISDAVSAVARA---ILPDES---ENEVIDLPERTALANKT 561  
DB 3215 DTIINGGASIEQITTEKIRBEATNALNEAKOHLTADTISLKTETVRKL--SRGDTNNK 3273  
QY 562 PADVYQSKKVPVLYVFVNASDKPRDQOIGLWGSDTGTRLVTFEHNILNRDGYOAGELRL 621  
DB 3274 PSSV-----SAYNNTIHSIQSELTQ 3293  
QY 622 SEDKKGVKLVATKPLS-----HPLNDQLRATLGYQOEYFGHSTNGFDLSTPTL 669  
DB 3294 TENNANTTI--NKPIRSVEEYNNALHEVNOQLNORLTDITINLOPL--ANKSELSKARRRL 3349  
QY 670 EHESTRSIIQNGGNRTYSLRYRLDKLTKQAPPETWODLPVDFVNGKRSQE 720  
DB 3350 ESKINET-VOTDGMTQOSVENTKQAKIKAOSSIAQTLL---INNGASDQ 3396  
RESULT 24

US-08-961-083-118  
;; Sequence 118, Application US/08961083  
;; Patent No. 6159469  
;; GENERAL INFORMATION:  
;; APPLICANT: Choi et. al.  
;; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
;; NUMBER OF SEQUENCES: 452  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Human Genome Sciences, Inc.  
;; STREET: 9410 Key West Avenue  
;; CITY: Rockville  
;; STATE: Maryland  
;; COUNTRY: USA  
;; ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 118:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1040 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-961-083-118  
Query Match 2.8%; Score 134; DB 4; Length 1040;  
Best Local Similarity 19.4%; Pred. No. 0.024;  
Matches 190; Conservative 124; Mismatches 318; Indels 348; Gaps 53;  
QY 24 MTSQALAOQNNPANI--INHP-----AHDTAIINQAKAGNPVLLTPPEIQARLNA 73  
DB 52 LTSQILSNVNSQLSGVGHLEPEPLKIEGYOYIGYIKTKKQNTLSRT---VDGKYSQ 108  
QY 74 GINAKPOQALDVNFD-----OSPISRIEQQSPPLGLDMSVIEETPLSLEELFAOE 127  
DB 109 RDSQPNSTKTSVYHSADELMNQGQKVSLOGEASGDGL-----SEKSIADNLSSND 163  
QY 128 STEMGINNDYIPEYQEQPNSEVVPPLEBEKPLIKRLVRLFNDGVNKPRLKAF 187  
DB 144 SFASQVEON--PDKKES---VVRPTV--PEQNPVSA----- 194  
QY 188 YOSQSGTSAIGSSHQKTEPYANIKALEDI--TQESAMDLNGSIPRLRQFALVARAVG 246  
DB 195 -TTVQSAEEVYATINDRE-----YKLPETKGTQEPGHE 229  
QY 247 YYDIDLSIRNSIGEVYIHDLEPVYIDYRAVEVRG--EGADKAFITVADEVPLLGD 305  
DB 230 -----GSA-AVREDL--PVYT--KPLETKGTQPGHGEAFAVREEPATYEP 271  
QY 306 VF-----HHGKETKKNLIE-----NABAHEGYFDGRMDSVYIILPDNAT 348  
DB 272 LATKGTQEPGHEKATVREETLETEPYATKGTQEPH---EG---EXVBEELP----- 320  
QY 349 VSLIYDTGYRFDVYFTTIDPKTNQLTDDPKLPVRKRELEQLLTVNMGAAYMLQAVR 408  
DB 321 -ALEVYTTNRRTETQNI-----PYTTEIQTDPITLKNRKRIRQ---GQA----- 360



SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 1326  
TYPE: PRT  
ORGANISM: Acetobacter xylinum  
FEATURE:  
OTHER INFORMATION: n at positions 15741 and 15767 may be a, g, t, or  
US-09-147-236-5

Query Match 2.8%; Score 130.5; DB 4; Length 1326;  
Best Local Similarity 19.6%; Pred. No. 0.072;  
Matches 197; Conservative 139; Mismatches 386; Indels 283; Gaps 52;

OY 47 TAINOAKGNPPVLT-PEQIOARL---MAAGLNKPOSQAL---DYVNDQSPIT-SR 97  
DB 181 TOMDARAGLGIYVSNQNYRAQALFYNTSTREMEGLTRKLDQSFQSQAPVEVA 240  
OY 98 IGEOSPGLGD-MSYIEETPLSELEFAOESTEMGPNNDIPEYQGEQPSSEVVPPT 156  
DB 241 AATOSYROTLSMLFVNPPTQPLMEQWLSA-----HPND-----AALREHMLHPBG 285  
OY 157 LEPEKPLIKRLYARLFENDGVNKKYPRKA--KFYQSS--SGETSALG-----SSHQ-- 204  
DB 286 GPPOKAGLARAQAGYQQLNAG-----RLSABEQSFQALQINSHDADSLQGMGLVSMROD 340  
OY 205 --KTEPVANIKAALEDITQE-----SAMDNGSIPLRQFALVA-----A 242  
DB 341 TAAHPIYEEAAMADPKYADRMKRALAGMAVSGDYAAVRQ--LIAHOYTEAKOKLATLA 398  
OY 243 RAYGVY--DIDLSTIRNSIGEVYIIDLGPYIDYRAVEVR-----GEGADDK 290  
DB 399 ROPQYGTATLMLADLQSTGCVAAEQ-----EYRGILSTREPNQALMGLARVUM 450  
OY 291 APTTVADEVPL-----IGDVFHKKYETKKNLENSAHEGYFDGWLDRSDV 340  
DB 451 AOGNTAEARQLLSRVSPQYASQVGEIEVSG-----LMAAASQ----- 487  
OY 341 ILPDTAAVSLIDTQGYRDEVEVFTIDPKTNLTDPKLPKRELEQLLTVNKE 400  
DB 488 -TSDARKVSTIREMAAPRDPWYRNL--ANALQOOGDAERGRVA--OPTLANPYT 541  
OY 401 AVNLQA---VRALSNDLIATRYFMNVTEIYFPEREQIQNDQVFEQSSSRTEPAQVD 456  
DB 542 AODRQAGLITYGSGNDAMTQQLAGLSPADSPAIRSLA--EEMELKODLASRLMVSMP 600  
OY 457 ESTLEPVIEVELDGIILDISPIEFSASNLIOCKLNLVAAKARHLVMDPRDV----- 510  
DB 601 VPLIREALSDPDPTGARVAVADLFROGRGDVNHAMALRIASTRTIDLSPPDRLSYATEY 660  
OY 511 LAINHDOCVNRSILGRISDAV--SAVARAILPDESE-----NEVIDLPERTALANKTPAD 564  
DB 661 MKISNPVAAAR-LIAPLDGDTGSGAGNMLPEQMOTLOQLRMGISVAOSDLINQGDQAO 719  
OY 565 VYQSKVPLVYFVASDKPR-----DGOIGLGWSDTGRVLTKEHNLINNDGYOA 615  
DB 720 AYDHLAPALQADPEATSPKLLARLYNGHKPKKALEID-----LAVLRHNPDDIDARQA 774  
OY 616 GAELRLSDKKGV--KLYATKPLSHPLNDQLRATIGYQOEYVGHSTNGFDSTRLEHET 673  
DB 775 AVQAAYVNSNHSLATRLAMDGVQESPMARAWLMAVADQADGH-----GORTIE-- 824  
OY 674 SRSIIIONGMRRTYSLARYLRLK-----KTQAPPET-----WQDLP 709  
DB 825 -----DLRRAYDL--RLQOVEGTRAASGPVGAHEELAPRSTINPQSRGYGHQVELG 874  
OY 710 VDFVNGKPSQA-----LLAGVA--VHKTVDNL--VNPMRGYRQRYSLVEGSSGL 756  
DB 875 APYVGYSYSAEASPDSTDOMLSIAQOIH--TLRENLAIPSIDGIGFSR--SGEHG-MGR 931  
OY 757 VSDANMALA-----RAGISGYVSPGDNAVGSNRHAQMTGGIOAGITWSDNF-----HVP 806  
DB 932 LTEANIPVCRPLQAGASA-----LTFSTPTMTWISQQLNTGVSVDVP 975

OY 807 YRLRFAGGDSIRGYANDSLSPIDKGYLTGG--QVLAVTAEXNYEPMKRLRLAVFGD 864  
DB 976 -RYXTFMATQANQACAGHSSCGGLD---FLSANHTQRIAGAAGE----- 1015  
OY 865 IGNAYDKGFTNDTKIGAGVGVNR-----ASPVQGVRVAVATGVK 903  
DB 1016 -----AGFAPDVORFNS-----WVRADVCASPIGPIITNVLGQVE 1050

RESULT 27  
US-08-323-477-2  
Sequence 2, Application us/08323477  
Patent No. 6086896

GENERAL INFORMATION:  
APPLICANT: Sparling, P. Frederick  
TITLE OF INVENTION: ANTIGENIC IRON REPRESSIBLE PROTEINS FROM  
TITLE OF INVENTION: N. MENINGITIDIS RELATED TO THE HEMOLYSIN FAMILY OF TOXINS  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Imclone Systems Incorporated  
STREET: 180 Varlick Street  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10014  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/323,477  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/920,963  
FILING DATE: 28-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Feil, Irving N.  
REGISTRATION NUMBER: 28,601  
REFERENCE/DOCKET NUMBER: SPA-2-2P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-645-1405  
TELEFAX: 212-645-2054  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1115 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-323-477-2

Query Match 2.7%; Score 129; DB 3; Length 1115;  
Best Local Similarity 19.9%; Pred. No. 0.072;  
Matches 205; Conservative 102; Mismatches 335; Indels 388; Gaps 45;

OY 56 NPVLLTPEQIOARLNAAGLNKPOSQALDVVNFDDQSPISIRGEQSPPLGLDMSYIEET 115  
DB 155 NSRMKSSSTVLOHLLVYLGFEKTK-----DYGN-----ERLGE-----SIMNIDPF 194  
OY 116 TPLSLEELFAOESTEMGPNNDIPEYQGEQPSSEVVPPTLEPEKPG----- 163  
DB 195 TPSKLANPFA-----DPTYYSNVLEVSFITYSLVDDANPMWKGGDPIYGRGISWNG 246  
OY 164 -LIKRLYARLFENDGVNKKYPRKAKEYQ---SSQGETSAIGSSHOKTEPYANIKALAD 218  
DB 247 ELLEKWKYKDF-----LPYLEKEWQDFPKFEDWLPPEPPENAREWMLKDPKRSKYHAYD 300  
OY 219 ITQESAMDNGSIPRLRQFALVAAR--ANGYDIDLSIIRNSIGEVY-----YIHDGEP 272  
DB 301 ---PLALDIDGD-----GIETVAAKGFAGALPDRNOGIRTAIGWVSAADDGLVRLNGN 352

```
QY 273 VTIDYRAVEYRGACDDKAFYVADEVPLLIGVFHHKYEYTKNLIENASAEHGYEDGR 332
      |||
      |||
      |||
Db 353 GIID-----NGAE-----LFGD-----NTK--LADGSFAKHGYAALA 382
      |||
      |||
      |||
QY 333 WIDRSYVILLPONTADYSL-----LYDTGTQRFDEYVFTID----- 370
      |||
      |||
      |||
Db 383 ELDSSNDNII--NAADAAFOTLRVWODLQDQISQANELFTLELGISDLAYKDVKN 440
      |||
      |||
      |||
QY 371 -----PKTNQLTTD-----PDKLPVKRELLBOLLVNMGEAY 402
      |||
      |||
      |||
Db 441 LGNGNTLAQGGSTYTKDGTAKMGDILLADNLSRKKDVELTAEQAKANLAGIGRLR 500
      |||
      |||
      |||
QY 403 NLQAVRALNDLIATRYFNNVNTETIYFPEREQIQNDQVSEQSSSRTEPAQVDESTLEP 462
      |||
      |||
      |||
Db 501 DLREAAALSGDLA-----NML-----KAVSAATKEAQL--ALLDN 534
      |||
      |||
      |||
QY 463 VIETVELTGLIMDISPIESA-----SNLIQDKLNLVAA--KARHLYDM 505
      |||
      |||
      |||
Db 535 LHKMAETDSNMGKSPMRSLSTDMQTQANEGIALTPSQVQLKMAVLSLSDAKAIDA 594
      |||
      |||
      |||
QY 506 PDDRVLAINHDDGVNRSILGRISDAVSAVARAILPDESENEVIDLPERTALANKTPADY 565
      |||
      |||
      |||
Db 595 ADRIVAVLAYTQDSSSTLYTMS-----EDALNIVKTYTNDT 631
      |||
      |||
      |||
QY 566 YOSKRVPLY--VFVASKPRDQIGLGWSDYGT--RLVTKFEH-----NLINR 610
      |||
      |||
      |||
Db 632 YDLAKNIYONLLFOTRLOPYLQNLQISPKMENDEFTLDFSGLYAFNNVYKFTNPQAKAVDL 691
      |||
      |||
      |||
QY 611 DGYQAGAEIR-----LSEDK----- 625
      |||
      |||
      |||
Db 692 AEMLAYGELSMYEGRRMLADYVEAKKGFEDYQKVLQETVALIALKTSQTQADDILQ 751
      |||
      |||
      |||
QY 626 -----KGVKLYATKPR--LSHPLNDOLRATLGYQOEVEGHS----- 658
      |||
      |||
      |||
Db 752 NNGFGINKNVSLKGNNDNTLIGAGNDYLEGSGSDTYTFGKFGQDITYVNTDYATGRK 811
      |||
      |||
      |||
QY 659 -----TNGF--DLSTRLE--HEISR-----SIIONGWNMTYSLRYRLDKLKT 698
      |||
      |||
      |||
Db 812 DIIREFDGTADMLFTTREGNHLLIKAKDSSGQVTVQSYQNDG-----SCAIRIDE--- 863
      |||
      |||
      |||
QY 699 QAPPEWQDLPVDFVNGKSSQEALLAGVAVHKTVADNLVPMRGYKQKRSLEVGS--SGL 756
      |||
      |||
      |||
Db 864 -----IHFDNGKYLDA-----YVKELYQOSTGDSRLYAYQSGSTLNGG 903
      |||
      |||
      |||
QY 757 VSDANMAIARAGISGYVSTGDNAVGSNRAHQMTGQIAGIYMSDNNHVRRLRFRAGCD 816
      |||
      |||
      |||
Db 904 LQD-----DLYYGADGNDLLGADAGNDSISGN-----GND 934
      |||
      |||
      |||
QY 817 QSIRGAHDSLSFISDKGYLTGGQVAVGTAEYNYEFMKDLRLAVFGDIGNAV--DKGFTN 875
      |||
      |||
      |||
Db 935 TLDDGGNDALVGYNGNDALNGE-----GNDHLNGEDGND--TLIGAGNDYLEGSGSGS 967
      |||
      |||
      |||
QY 876 DTKI--GAGVG 884
      |||
      |||
      |||
Db 988 DTYVEGEGFG 997
      |||
      |||
      |||
RESULT 28
US-09-134-001C-4547
; Sequence 4547, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
```

```
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4547
; LENGTH: 675
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (18)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-134-001C-4547
```

```
Query Match 2.7%; Score 128.5; DB 4; Length 675;
Best Local Similarity 18.6%; Pred. No. 0.034;
Matches 152; Conservative 96; Mismatches 267; Indels 303; Gaps 43;
```

```
QY 187 FYOSOSGETSAIGSSHQKTEPANTAKALEDITQESAMDLNGSIPLRQTALVAARAVG 246
      |||
      |||
      |||
Db 29 FYASKDKEINNTIDALEDK-----NFKQVYKSSYSKSD--NGEVE-----MTERPIK 75
      |||
      |||
      |||
QY 247 YPIDLSIIRNSIGENVDTIHDLGEPYIDYRAVEYRGACDDKAFYVADEVPLLIGDV 306
      |||
      |||
      |||
Db 76 IY-----NSLGVADINIQDR-----KIKVSKKKKRYD----- 103
      |||
      |||
      |||
QY 307 FHHGKYEYTKR--NLIENASAHEGYEDGRW--LDRSYVDVILLPDMTADYSLIYDTGTQYRFD 362
      |||
      |||
      |||
Db 104 ---AQYKIKTNVGNIDRNVQFNFYKEDGMKLDMDHVSIIIPGMQDOSI----- 149
      |||
      |||
      |||
QY 363 EYVEFTIDPPTNQLTDPDKLPYKRELLBOLLVNMGEAYNLQAV-----RALSDN 413
      |||
      |||
      |||
Db 150 -----HIENLKSERGI--LDRNVE--LANGTAYEIGIVPKNVSKKDYKALAKE 196
      |||
      |||
      |||
QY 414 LIATRYFNNVNTETIYFPEREQIQNDQVSEQSSSRTEPAQVDESTLEPIETVELTDLGI 473
      |||
      |||
      |||
Db 197 L-----SISEDYIKQOMQON-----WVQDDTFRP--LKYVKKKDEY 230
      |||
      |||
      |||
QY 474 LMDIS-----PIEFSASNLI--ODKLNVAKAKRHLMDPDRVLA----- 512
      |||
      |||
      |||
Db 231 LSDFAKFFHLTTNETESRNPPLGKATSHLGVYGPRIINSEELKQKEYGYKDDAVIGKGL 290
      |||
      |||
      |||
QY 513 -----INHDDGVNRSILGRISDAVSAVARAILPDESENEVIDLPERTALANKTPADY 565
      |||
      |||
      |||
Db 291 EKLDRKKLQHEHDYRYTI---VDNSNTIAHTLI---EKKKKDKXIOULTIDAKYQKSI 343
      |||
      |||
      |||
QY 566 YOSKRVPLYVFVASKPRDQIGLGWSDYGTRL-----VTKREHNLINRDGYQAG 616
      |||
      |||
      |||
Db 344 YNNMK-----NDYSGSTAIHPQTGELLALVSTPSYDVYPPMYGMSNE--YN-- 388
      |||
      |||
      |||
QY 617 AELRLSEDKKGVRLYATKPLSHPLND--QLRATLGYOQOEVEGHSTNGFDLSTRLEHEISR 675
      |||
      |||
      |||
Db 389 ---KLTEDEK-----EPL---LNKFQITTSFGSQKIL---TAMIGLNKKTLDDKTSY 432
      |||
      |||
      |||
QY 676 SIIONGWNMTYSLRYRLDKLKTQAPPEWQDLPV--DFVNGKPSQEALLAGVAVHKTV 732
      |||
      |||
      |||
Db 433 K-IDGKGWQK-----DK-----SWGGYNVTRYEVVNGNIDILKQAIES----- 468
      |||
      |||
      |||
QY 733 ADNLVPMRGYRQKRSLEVSS-----GLVSDA-----NMAIAAGISGYVSEF 776
      |||
      |||
      |||
Db 469 SDNI-----FPAVALDELGSKKFEKGMKLLGVGEDIPSDYPPYMAOISKNLNDNETLLA 522
      |||
      |||
      |||
QY 777 DNAYGSNR-----AHQMTGQIAGYIWSDFNHNV-----PYRLRFFAGD 816
      |||
      |||
      |||
Db 523 DSGYGGGELLINVOILSTISALENNGNINARHLLADTKKVKKKNIKENINLTLTDM 562
      |||
      |||
      |||
QY 817 QSI-----RGYAH-----DSLSPISDKGYLTGGQV-----LAV----- 844
      |||
      |||
      |||
Db 583 QQVYVNTKHEDIYRSYANLIGKSGTAELKKQKQETQKQIOWFTISYDKDNPMMAALNVKD 642
      |||
      |||
      |||
QY 845 ---GTAEYNYEFMKDLRLAVFGDIGNAVYDKGFTNTYK 878
      |||
      |||
      |||
Db 643 VQDKGMASTNAKIS-----GKYDELXENGNK 669
      |||
      |||
      |||
RESULT 29
```

US-08-962-859A-2  
Sequence 2, Application US/08962859A  
Patent No. 6127345  
GENERAL INFORMATION:  
APPLICANT: Burnham, Martin K. R.  
TITLE OF INVENTION: No. 6127345el Glucose 6-Phosphate  
Dehydrogenase Gene  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Dechert, Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2793  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compulide  
OPERATING SYSTEM: Windows  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/962,859A  
FILING DATE: 03-NOV-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/035,072  
FILING DATE: 21-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson, Arthur E  
REGISTRATION NUMBER: 34,354  
REFERENCE/DOCKET NUMBER: GM50001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 609/520-3254  
TELEFAX: 609/520-3259  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 495 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-962-859A-2

Query Match 2.7%; Score 127.5; DB 3; Length 495;  
Best Local Similarity 19.9%; Pred. No. 0.024;  
Matches 115; Conservative 93; Mismatches 202; Indels 169; Gaps 32;

QY 148 NSEVVVPTLEPEKPLIKR-LYARLFNDGVNKKVPRKAKFYOSSGSETSAIGSSHQKT 206  
DB 2 SSKVIV--TIFGASGLDAKKRLY-----PSLLDLKXSGNLSKHFVIGTARR-- 46  
QY 207 EPPANIKALEDITOESAMDLNGSIPRLKOTALVAARAVGYDIDLSIRNSICEVDYII 266  
DB 47 -PWS--KEYESVYVESILDLADSTEOAQEF--SHEFYQ-----S 82  
QY 267 HDGGE-PVYIDYRAVEY---RGEADDKAFTVADEVPLLIGDVFNHGKYEKTKNLLEN 321  
DB 83 HDVNDLEHTALAKLOALENEKYAENKKLF--FLSMAPQFFGTIAHKSE---NIV-- 135  
QY 322 ASAEHGVEFDGRWLDKRVNLDVLPNTADVSLIYDTGTQYREDEVVFTIDPKTNOLTTDPD 381  
DB 136 -----DCKGFRLI-VEKPFCT-----DYATASKLMDLAFDE----- 169  
QY 382 KLPVKRRLLEQLTLVN--MGEAVMLQAVRALSNLDLITRTFNNMNTLEVFPEREQIONDO 439  
DB 170 -----BOIFRIDHYLG-----KEMIQSIFAVREFANLLFENVW--NKDFINDR 210  
QY 440 VSEFOSSSSPTPEAQQVES-TLEPVIE--TVELFDGLIMDISPIEFSASNLIOCKLNLVA 496  
DB 211 ITTAERIGVYERGGYDQSCALROMVGNHTLIQLISLAMD-KPASFTKDELRAEKIYF- 268  
QY 497 AKARHLVMDPDRLVALINHDGCVNRSTLGRISDAVSAVARAILDESENEVIDLPETAL 556

DB 269 ---KNLYH-PTDEELKEHFRGOYRS--GKI-DGMYIYSTRSEPN----- 306  
QY 557 ANRKTADYQSKKVPPLYVEVAVSDKPRDGOIGLGWSDGTGLVTKFEH-NLINRDYQA 615  
DB 307 VNPESTTEFTSGA-----FVYDSRFR-----GVPFRTGKRLTEKGTNIVY----- 351  
QY 616 GAELRLSEDKKGVKLYATKPLSHPLNDQLRATLGYQOEVEGHSINGFDLSRTLEHLSR 675  
DB 352 -----FKQWDSIFGEPLAPNI---LTIYIQPEGFSLSINGQVGEEF 391  
QY 676 SIIONGGMNRTYSRLRYRLDKLTOAPPEMODLPVDFVN 714  
DB 392 NLAPN-----SLDYRTDATATGASPEPYEKLITVDLN 423

RESULT 30  
US-09-134-001C-5584  
Sequence 5584, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 5584  
LENGTH: 801  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5584

Query Match 2.6%; Score 125; DB 4; Length 801;  
Best Local Similarity 18.1%; Pred. No. 0.09;  
Matches 127; Conservative 115; Mismatches 239; Indels 222; Gaps 29;

QY 119 SLEELFAQESTE-----MGINPNDYIPEYQGEQPNSEVVVPTLEPEKPLIKRL 168  
DB 15 SIEMINQOPVEPMSVSDFODALGINSADSFDR-----LIKILVLEQSLIE-- 62  
QY 169 YARLFNDGVNKKVPRKAKFYOSSGSETSA---IGSSHQKTEPPANIKALEDITOESAM 225  
DB 63 -----RTRIDRYQRQSNKTSKLIKIGTJSQNKKGAFRL-----PEDEM 103  
QY 226 DLNGSIPRLKOTALVAARAVGYDIDLSI--IRNSICEVDYIIDLGPYIDYRAV-EV 282  
DB 104 DDITPTPKIKRAL-----DGDYIYELQKSRGHNKKIE--GVYKSIENHSYQV 152  
QY 283 RGEADDKAFTVADEVPLLIGDV---HNKGYEKTKNLI-----ENASAEHGY 328  
DB 153 VGTYSSEAKHFGFVLPDDKRIMQDIFIRKGMIGAVDGHKVLVQITKYADSTDNPEGHVA 212  
QY 329 FDGWRWLDKRVNLDVLPNTADVSLIYDTGTQYREDEVVFTIDPKTNOLTTDPDKLVKRE 388  
DB 213 ILGHRNDPBGVDIL-----SIYQHGIELEFPDVLQAEAEVPDVI--EPSELEGRRD 262  
QY 389 LLEQL-LTVNMGAEVYNLQAVRALSNLDLITRTFNNMNTLEIYFPEREQIONDOVSFEQSSS 447  
DB 263 LRDELITIIDGADAKDL-----DDAIVAVKKLKNGNELI-----VSINDVSYVWEG 309  
QY 448 SRTEPAQVESTLEPIETV-----E 468  
DB 310 SALDKRAVATRSYLVLDVRIPIHRLSNGISCLNPEEDRLTISCRMEINEREVEVYKHE 369  
QY 469 LTDDGL-----MDISPIEFSASNLIOCKLNLVAKARHL 502  
DB 370 IFDSVISHSNMYMTYDAVANKIITDDQSEIRSOYKDLTPMLDLAODLSNRLIMRRKRGEID 429



[illegible]

RESULT 31  
US-08-537

```

? Sequence 3, Application US/08537002A
? Patent No. 5773282
? GENERAL INFORMATION:
? APPLICANT: TSUSAKI, Keiji
? APPLICANT: KUBOTA, Michio
? APPLICANT: SUGIMOTO, Toshiyuki
? TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
?   TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
? NUMBER OF SEQUENCES: 17
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: BROWDY AND NEIMARK
? STREET: 419 Seventh Street, N.W., Suite 300
? CITY: Washington
? STATE: D.C.
? COUNTRY: USA
? ZIP: 20004
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/537,002A
? FILING DATE: 29-SEP-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 260984/1994
? FILING DATE: 01-OCT-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP No. 5773282 yet received
? FILING DATE: 08-SEP-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: BROWDY, Roger L.
? REGISTRATION NUMBER: 25,618
? REFERENCE/DOCKET NUMBER: TSUSAKI-1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-628-5197
? TELEFAX: 202-737-3528
? TELEX: 248633
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 963 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? IS-08-537-002A-3

```

```

Query Match      2.68;   Score 125;   DB 1;   Length 963;
Best Local Similarity 24.38;   Pred. No. 0.12;
Matches 98;   Conservative 46;   Mismatches 130;   Indels 130;   Gaps 23;

4 PV-LFANRSEFMPALAAVLPIMTSQALQOONNPANINHHVPAHDTAINQA-----52
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db      507 PVELFSQGFPEVPEGRYRLTGLGPHGFALFKPVEAVLHLRSPDMAEPPAEADLRVH 566
Qy      53 KAGNPVLLTPROIOAR-----LNA-----AGLNKKPOSQA-LDVVNFDDOSPISRI 98
Db      567 MGGGEVLLVDVLIVERNGEREELLNAOTLEKSWLAKPOKVALLDALRF-----6171
Qy      99 GEOSPPLGLDMSVIE--ETTPSLLEELAOEST-----EMGINPDY- 138
Db      618 -QKDPRLYITTLQLENNHTLOSLEPLMSPORREGPGLEAFTHGQPGFYELSLDPGEYR 676
Qy      139 -----IPE-----YQGEOPNSEVVNPPTLEPEKPGILKR-----LVARLENDQ--- 176
Db      677 LLARLKEGFEGRSRLRAYYRGHNER---VEAVDILRLPGLAGEGVMYQGLVQDDGLD 733
Qy      177 ---VKNVPRLK-----AKFYOSSOGETSAIGSSHOKTEPYANIKALEDITTOESAM 225
Db      734 RTEVRLPRLDLPFWLPRGEGFEMWEGASRYVALTGLSLPGRP-QDLFAALEVRLLESIP 792
Qy      226 DLNGSLP-----RLNGT-ALVAARAVGYDDLDLSIRNSIGEVYII--HDL-----268
Db      793 RLRGHAPGTPGLLPGALHETALV--RLLG--VRLALLRLALEVGVGVGHPLLRGRL 847
Qy      270 -----GEPVYIDYRAVEVEGEGADKAFETTVADENVPLILGDV 306
Db      848 GAFLELEGE-VYIVALGAEKRG-----TVEEDLARLAYDV 881

```

RESULT 3  
US-08-86

Sequence 3, Application US/08863010  
Patent No. 6087146  
GENERAL INFORMATION:  
APPLICANT: TSUSAKI, Keiji  
APPLICANT: KUBOTA, Michio  
APPLICANT: SUGIMOTO, Toshiyuki  
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR  
TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/863,010  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/537,002  
FILING DATE: 29-SEP-1995  
APPLICATION NUMBER: JP 260984/1994  
FILING DATE: 01-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP No. 6087146 yet received  
FILING DATE: 08-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TSUSAKI-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 963 amino acids

```

: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-863-010-3

Query Match
Best Local Similarity 2.6%; Score 125; DB 3; Length 963;
Matches 98; Conservative 46; Mismatches 130; Indels 130; Gaps 23;

OY 4 PV-LFANRSMFVALAAYLPLMTSQAALAOQNNPANIINHVPADTAIQA-----52
DB 507 PVLEFSQGFPPVEGGRYRLTLGPHGFALFKVEAVLHLPSPDMAEERPAPEADLPVH 566
OY 53 KAGNPVLLTPRQIOAR-----LNA-----AGLNARQSOA-LDVVNFDDQSPISRI 98
DB 567 MGGPEVLLVDTLVHGRGRELNLALQTLKESWLAKPKQVALLDALRF-----617
OY 99 GOSPPGLGDMSVIE--ETTPLSLELFAQEST-----EMGINPNDY-138
DB 618 -QKDPRLVTLTLQLENHRTLOVSLPLMSPOREBGLFARTHGOGFYELSLDGFYR 676
OY 139 -----IPE-----YQEQPNSEVVVPPTLEPEKPKLIKR---LYARLFNDG---176
DB 677 LLARLKEGFEGRSRLRAYYRGHRNPGP---VPRAVDLLRGLAAGEVWQGLVODGGLD 733
OY 177 --VNKVPRLK-----AKFYQSSQSGETSAIGSSHOKTEPYANIKALEDTIOESAM 225
DB 734 RTERVLPRLDLPWVLRPEBGLFMEGASRRVLTGSLPPGRP-ODLFALEVRLLIESLP 792
OY 226 DLNGSIP-----RLROT-ALVARAVGYDDIDLSIIRNSIGEVYII--HDL-----269
DB 793 RLRGHAPGTGGLPGALHETALV--RLLG---VRLALLHRLALGEEVGVGHPILGRGL 847
OY 270 -----GEPVYIDYRAVEVRGEGADKAFITVADEVPLIGDV 306
DB 848 GAFLEEGE-VYLVALGAEKRG-----TVEEDLARLAVDY 881

RESULT 33
US-09-024-429-3
: Sequence 3, Application US/09024429
: Patent No. 6165768
: GENERAL INFORMATION:
: APPLICANT: TSUSAKI, Keiji
: APPLICANT: KUBOTA, Michio
: APPLICANT: SUCIMOTO, Toshiyuki
: TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
: TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/024,429
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/537,002
: FILING DATE: 29-SEP-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 260984/1994
: FILING DATE: 01-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 255829/1995
```

```

: FILING DATE: 08-SEP-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: YUN, Allen C.
: REGISTRATION NUMBER: 37,971
: REFERENCE/DOCKET NUMBER: TSUSAKI-1B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: TELEX: 248633
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 963 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-09-024-429-3

Query Match
Best Local Similarity 2.6%; Score 125; DB 4; Length 963;
Matches 98; Conservative 46; Mismatches 130; Indels 130; Gaps 23;

OY 4 PV-LFANRSMFVALAAYLPLMTSQAALAOQNNPANIINHVPADTAIQA-----52
DB 507 PVLEFSQGFPPVEGGRYRLTLGPHGFALFKVEAVLHLPSPDMAEERPAPEADLPVH 566
OY 53 KAGNPVLLTPRQIOAR-----LNA-----AGLNARQSOA-LDVVNFDDQSPISRI 98
DB 567 MGGPEVLLVDTLVHGRGRELNLALQTLKESWLAKPKQVALLDALRF-----617
OY 99 GOSPPGLGDMSVIE--ETTPLSLELFAQEST-----EMGINPNDY-138
DB 618 -QKDPRLVTLTLQLENHRTLOVSLPLMSPOREBGLFARTHGOGFYELSLDGFYR 676
OY 139 -----IPE-----YQEQPNSEVVVPPTLEPEKPKLIKR---LYARLFNDG---176
DB 677 LLARLKEGFEGRSRLRAYYRGHRNPGP---VPRAVDLLRGLAAGEVWQGLVODGGLD 733
OY 177 --VNKVPRLK-----AKFYQSSQSGETSAIGSSHOKTEPYANIKALEDTIOESAM 225
DB 734 RTERVLPRLDLPWVLRPEBGLFMEGASRRVLTGSLPPGRP-ODLFALEVRLLIESLP 792
OY 226 DLNGSIP-----RLROT-ALVARAVGYDDIDLSIIRNSIGEVYII--HDL-----269
DB 793 RLRGHAPGTGGLPGALHETALV--RLLG---VRLALLHRLALGEEVGVGHPILGRGL 847
OY 270 -----GEPVYIDYRAVEVRGEGADKAFITVADEVPLIGDV 306
DB 848 GAFLEEGE-VYLVALGAEKRG-----TVEEDLARLAVDY 881

RESULT 34
US-09-134-001C-3172
: Sequence 3172, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3172
: LENGTH: 563
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-3172
```



Db 780 KDELGS 786

RESULT 36  
US-08-719-822B-2  
Sequence 2, Application US/08719822B  
Patent No. 5874527  
GENERAL INFORMATION:  
APPLICANT: Barnwell, John  
TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby and Darby  
STREET: 805 Third Ave.  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022-7513  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/719,822B  
FILING DATE: 09/30/96  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 5986/17686052  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)527-7700  
TELEFAX: (212)753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1018 amino acids  
TYPE: amino acid  
MOLECULE TYPE: linear  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
FRAGMENT TYPE: C-terminal  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium vivax  
IMMEDIATE SOURCE:  
CLONE: PVMB3.3.1  
US-08-719-822B-2

Query Match 2.68; Score 124; DB 2; Length 1018;  
Best Local Similarity 18.6%; Pred. No. 0.16;  
Matches 124; Conservative 100; Mismatches 227; Indels 216; Gaps 31:

QY 30 AQON-NPAN-----IINVPADDTLINAQAGN-----PPVLLTP 63  
DB 230 AQENVOPADSDALFGSLDKDIEFDHKPELFQIVAGTAKHVGTGOELPMKPVLPV 289  
QY 64 EQLQARLNAAGLNAQPSO--ALDVVNFDDQSPRIGSOPRLGDMKVTEETPLSLSE 122  
DB 290 AEEQAQVPAEFLDTPDDFDALDT-----ESP-----BEVELVDEE 327  
QY 123 LFAQESTKMCINPDYIPREYQGEQPNSEVVVPTLEPEKPLIRLYARLFNDG--VNVK 180  
DB 328 ATEESTEVG--PTE-----EGPTEELDTP-----EDGFRIRRN 360  
QY 181 PRKAKFYSSGGESFSAISSHOKT-EPAANKAALEDITQESAMDLNGSIPRLQOTAL 239  
DB 361 CRRNRNNEGEEETEAEAEVGEVPEEGEELEAPED--DFALD-GTTLLEETETAE 416  
QY 240 VAARAVGYDIDESTIRNSIGEVDVLIHDGE-----PVIYDRAVVRGEGADDKAF 292

Db 417 GEETVEGEETVE-----GEETVEGEAEAGPEELEATPEDDFQLEEPSGE----- 463  
QY 293 TTVADVEPLLIGVFHNGKYETKKNLIENASAEHGFDGRMLDRSDVILPDN----- 345  
DB 464 -----GEGEGEGEGEALVAVPVAA-----PVEVYTPAQVPRMVA 501  
QY 346 -TADVSLIYD-TGTQYGFDEVFPTIDPKTNQLTTPD-----KLPAKRELBGLLVNM 398  
DB 502 PTADETLFVILNDLTYADITSF--EPLRKQILKDDPAGEAVTVPSKAPVOVPVAVGP 559  
QY 399 GEAVNIQAVRALNSDLIATRYPMKAVNT-----EIVF-----PEREQIQDVS 441  
DB 560 AQEVPTEELMQLQED-----DFLEGTAAPAEGBELVLEBGEPTEREPREGEPTBEV 614  
QY 442 FROSSSRTEPAQVDESTLPVIEIV-----ELTDGLIMDS-----PIEFSASNLI 488  
DB 615 EELELATPEDDFELEPTGEVEVEGEETABGEVEEVPAAVEEVEEVPAAVEEVEEV 674  
QY 489 QDKLNTVAARARHLYDPDPRVLAINHDDGVNRSILGRISDAVSAVARAI--LPDESEN- 545  
DB 675 PEEVEEVPAAVEEVEEVEPEE-----VEEVEEVEEVEEVEEVEEVEEVEEVEE 719  
QY 546 ---EVIDLP-----ERTALANRRTPADVYOSKR-----VPLVFEVASDKPR 583  
DB 720 EEEVEEVPAAVEEVPAAVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 779  
QY 584 DQOIGIG 590  
DB 780 KDELGS 786

RESULT 37  
US-09-092-458-2  
Sequence 2, Application US/09092458  
Patent No. 6231861  
GENERAL INFORMATION:  
APPLICANT: Barnwell, John  
TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,  
TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby and Darby  
STREET: 805 Third Ave.  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022-7513  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/092,458  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/719,821  
FILING DATE: 09/30/96  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 5986/17686053  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)527-7700  
TELEFAX: (212)753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1018 amino acids  
TYPE: amino acid  
MOLECULE TYPE: linear

HYPOHETICAL: YES  
ANTI-SENSE: NO  
FRAGMENT TYPE: C-terminal  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium vivax  
IMMEDIATE SOURCE:  
CLONE: pVMB3.3.1  
US-09-092-458-2

Query Match 2.6%; Score 124; DB 4; Length 1018;  
Best Local Similarity 18.6%; Pred. No. 0.16;  
Matches 124; Conservative 100; Mismatches 227; Indels 216; Gaps 31;

QY 30 AQON-NPAN-----IINHPADRAINQAKGN-----PPVLLTP 63  
DB 230 AQENVOPADSNDALFGSILDKDILFDHIDKDFEPLFEQIVAGTAKHYTGQELPKMPVPLPV 289  
QY 64 EQQARLNMAGLNAKPOQO-AADVNVNDDQSPISRIGQSPPLGLMSVIEETTPLSLEE 122  
DB 290 AEPRAVPPEELDAPDPPDLDVLT-----ESP-----EEVELVLDDE 327  
QY 123 LFAOESTEMGINPDYIPEYQGEOPNSEVVVPTLEPEKPGLIKRLVRLFNQ--VNKV 180  
DB 328 ATEESTEVG--PTE-----EGPTLELATP-----EDGFRIRRN 360  
QY 181 PRKAKFYOSSQSGETSAIGSSHOQT-EPYANIKALEDITOSAMDUNGSIPLROTAL 239  
DB 361 CRRNRNRNVEGETEEEAEGVESEETPEGELEATPED--DEALD-GTTLEETETAE 416  
QY 240 VAAVANGYDIDLSIRNSIGEVDYIHDGE-----FYIDRYAVEYRGAGADKAF 292  
DB 417 GEETVEGETVE-----GAETVEGEAAEGEELEATPEDDPQLEPSEGE----- 463  
QY 293 TVVADEVPILGDFVHNGKETKKNLIENASAEHGYFDGRWLDRSYVLLPDN----- 345  
DB 464 -----GEGEGEGEGEGALVAVYVAE-----PVEVTPAQPKPMA 501  
QY 346 -TADVSLIYD-TGTQYRFEVVEFTTIDPKTNQITLTPD-----KLPRKRELLQLLTVNM 398  
DB 502 PRADETLFVDILNDLTYADITSF--EPLFKQILKDPDAGEAVTVPSKEAPVQVAVVCP 559  
QY 399 GEAYNQAARALSNOLIAIRYFMVMT-----EYVF-----PREDIQNDQVS 441  
DB 560 AOEVPTEELMOLED-----DELEGTAAPEBEGELVLEBEGPTEPREGEGTECEVP 614  
QY 442 FPOSSSSRTEPAQVDESTEPIVETV-----ELTDGILMDS-----PIEFSASNLI 488  
DB 615 EBELEATPEDELEPTEGEEVETVGEETAGEVEVEVPAVEVEVEVPAVEVEVEV 674  
QY 489 ODKLNLVAAKARHLXDPDDRYLAINHDDGVNRSILGRISDAVSAVARAI--LPDESEN- 545  
DB 675 PREVEVPAAVEVEVEVEVEE-----VEEYVEVEVEVEVEVEVEVEVEVEV 719  
QY 546 ---EVIDLP-----ERTALANKRTPADVYQSK-----VPLYVVASDKPR 563  
DB 720 EEVEVEVPAVAVEVEVPAVEVEVEVEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 779  
QY 584 DGOIGLG 590  
DB 780 KDELGSG 786

RESULT 38  
US-08-769-309A-5  
Sequence 5, Application US/08769309A  
Patent No. 5741890  
GENERAL INFORMATION:  
APPLICANT: Scott, John D.,  
APPLICANT: Nauert, Brian J.,  
APPLICANT: Klauack, Theresa M.  
TITLE OF INVENTION: Protein Binding Domains of Gravin  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower/233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,309A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5741890and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/33451  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEO ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1780 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-769-309A-5

Query Match 2.6%; Score 124; DB 1; Length 1780;  
Best Local Similarity 19.2%; Pred. No. 0.42;  
Matches 171; Conservative 129; Mismatches 294; Indels 298; Gaps 44;

QY 23 LMTSQLAQ-----ONPANIINHPAH-----DTAINQAKAGNPVLLTPEQIQRINAA 73  
DB 930 LTFEEVLEREVIAEEBPPTTEPLEPENREARQDIVSEAE-----LTPAVTAALTAG 982  
QY 74 GLNAKPOQALDVNFPDQ--SPISRIGQSPPLGLMSVIEETTP----- 118  
DB 983 PLSSEGTESAEEETTEMVSAVSQILD--SP-----DTTEATPPQVEGEGVPIDEQE 1035  
QY 119 -----SLEELFAOESTMGI--NPNDYIPEYO--GEOPNSVVVPTLEPEKPKLIK 166  
DB 1036 RRIQEVLOAAVEYKESQPLGTGEPDVLQPVORAEAPREDA-----EASGLK 1087  
QY 167 RLVARLNDGVNRPRLKAK---FYOSSQGETS-----AIGSSHOQT-----E 207  
DB 1088 E-----TDVVLKVDAGQEAETFTQGVGOTTPESEFKAPOVTESESSELVTTQAE 1141  
QY 208 PYANIKAALEDITOSAM-----DLNGSIPLROTALVAAVANGYDIDLSIR 256  
DB 1142 TLGVKS--QEWMEQALIPDSVETPTDSETPDSTP-----VADFPAPGTOK 1187  
QY 257 NSIGEDVYIHDGEVYIDYRAVEYRGEGADKAPTVADEVPLIGDFVHNGKETK 316  
DB 1188 DELVE-----IHEENE-----VHLVPVPGTEAE--AVPAOKERPAPSPVFOE--ETKE 1233  
QY 317 NLLENASAEHGYFDGRWLDRSYVDILPDMTADVSLIYDT--GTO-----YRDEVEVFF- 367  
DB 1234 QSKMEDTLEH-----TDKEVSV-----EYVSIISKTEGTQDAQYADEKTKDVPFEE 1280  
QY 368 ---TIDPKTNQITLTPDKL-----PVRKRELE 391  
DB 1281 GLEGSID---TGITVSRKYTEVALKGEETAECKKDALBELQSHAKSPSPVVEREMVY 1337  
QY 392 QL-----LVNNGEAYNQAARALSNOLIAIRYFMVMTVEIYFPERE----- 433  
DB 1338 OVEREKTEAPPHVNEKEKHEHTAVTV--SEEVSKQLQTVNVPIDIGAKEVSSLGSSP 1395  
QY 434 -----QIQNDVSEFOSSSRTEPAQVVESTLEPIETVELLDGILMDSPI 480



```

: COUNTRY: USA
: ZIP: 92121
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/436,664
: FILING DATE: 08-MAY-1995
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/394,232
: FILING DATE:
: APPLICATION NUMBER: 08/307,410
: FILING DATE: 16-SEP-1994
: APPLICATION NUMBER: 08/222,612
: FILING DATE: 16-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Fisher, Carlos A
: REGISTRATION NUMBER: 36,510
: REFERENCE/DOCKET NUMBER: GP94003.CP2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-535-2807
: TELEFAX: 619-546-7929
: TELEX:
: INFORMATION FOR SEQ ID NO: 34:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 876 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: Internal
: ORIGINAL SOURCE:
: US-08-436-664-34

```

```

Query Match      2.6%; Score 123.5; DB 2; Length 876;
Best Local Similarity 19.2%; Pred. No. 0.14;
Matches 193; Conservative 147; Mismatches 358; Indels 307; Gaps 51;

```

```

QY 8 ANSEMPVAL-----AAY-LPLMTSQAQAQNNPANIINHPA-----HDTAIN 50
DB 14 AYRAFALPLHNDKGIHTNAVYGFTMLNKILAE-QPTHILVAFDAGKTFRRHER-FQ 71
QY 51 QAKAG---NPPVL-----LTPBOIQAALNAAAGLNAKPOQALDVNFDOSPISRIGESQ 102
DB 72 DAKGGHQGTPELSEQFPPLRELLKAV-----RIPAYELDHYEADDIIGTMAARA 121
QY 103 PGLGMSV-----EETTPLSLEELFAQESTEMGINP-NDYIPEQGEQRPSEVVPPTLE 158
DB 122 ERGEGFAVKYISGRDQLTQLASPOV-TVEITKKGITDIESTYPE-----TVEKYGLT 172
QY 159 PEKPGILKRLYALFNDGVNKPRLAKAFYQSSQSGTSAIGSSHOKTEPYANIKALED 218
DB 173 PEQIVDLKLGMDK-SDNIPGVPGI-----GEKTAV-----KLLKQFGIVENVLAS 217
QY 219 ITQESAMDLNGSLPRLQRLALVAAR-AGVYYDIDLSTIIRNSI---GE-----VDYIHDLG 270
DB 218 IDEIKGEKLEKRNROYRDIALLSKQLAALICRDAPEVELTLDIYKGGEDREKVVALFQELG 277
QY 271 EPIYIDYRAVEV-RGE-----GADDKAFTTVADE-----VPLLIGDVFHNGKYEETKKN 317
DB 278 FOSFLKMAVOTDEGEKPLAGMDFAIADSVTDLADKALVAIVEVGDNYHHAP----- 331
QY 318 LIENASA-EGHYFDGRWLDRSVVILPDNT-ADVSLYIDGTQYREDEVVFTI-DEPTN 374
DB 332 IYGIALANERGR-----FLRPETALADPKFLAWLG-----DETCKKTMFDSKRA 376
QY 375 QLTTPDKLPVKRELLLEOLLTVNMGEAYNLQAVRAISNDLIAFRYENMVTETIVPEREQ 434

```

```

DB 377 AVALKMKGIELRGVVEDLLA-----AYLDDPAQA-AGDVAAMAKM-----QYEA 421
QY 435 IONDOVFEOSSSSRTPEACVDESLPEVIEFTVELTDLGILMDISPIESASNLIQDKLNL 494
DB 422 VRSDAVYKK-GAKRTVP--DEPTL----- 443
QY 495 VAAKARHLYDMPDDRVALINHDDGVNRSILGRISDAVASAVARAILPESENE-----VI 548
DB 444 ---AEH-----LARKAAAIWALEEPLIMDELRRNEQRLLTTEL 477
QY 549 DLPERTALANKRTPADVYQSKVPLVYFVASDKRRDQIGIGMKSDDTGLTVTFEHLI 608
DB 478 EQLAGILANNE-----FTGVKVDTRLEDM-----GAELTEQLQ--AV 514
QY 609 NRDCYQ-AGAEPLREDEKKGKLYATKPLSHPLNDOLRATLGYOEYFGHSTNGFDLSTR 667
DB 515 ERRIYELAGOEENINSRKQ-----LGVTLFDKL-----QLPYLKKTKTGTSTISAD 559
QY 668 TLE-----HEISRSII---ONGGWNRTY-----SLRYRLDKLTKQAPPET----- 704
DB 560 VLEKLAPHHEIYEIILHYRQLGKLGSTYIEGLLKVVHPVGTGVHTMFNOALTGTGRLSV 619
QY 705 ---WQDLPVDFVNGKRPQOEALLAG-----VAVHKTYADNLVPMNGYR 744
DB 620 EPNLQNIPIRLEEGRKIRQAFVPSPPMLIFPADYSOIETELVLAHIEDDNLIEAFRGL 679
QY 745 QRYSELEVSSGLVSD---ANMAIARAGISGVSEFGDNAGVSNRAHOMTGTIOAGYIMSD 800
DB 680 DIHTKTMDIFHSEEDVTANMRQAKAVNFGIYIGISDGLAONLITTKREAEPF--- 736
QY 801 NFNHVPYRLRFFAGDOSIRGYAHSLSPISDKGYLTGGVLAAGTAEYNYEFKDLRLA 860
DB 737 -----ERYFA-SFPGVQYQMDNIYQEAQKQGYV-----TLHRRRYLPDITSR 779
QY 861 VFEGDIGANAYDKGFTNDTKIGAGVGVNMAASPYGVYRVDAVATGVKEE 905
DB 780 NFN-----VRSFAERTAMNTPIOGSAADIKKAMIDLVSFLREE 818

```

```

Search completed: April 28, 2003, 16:24:54
Job time : 65 secs

```







```
OY 275 IDYRA-VEVRGEGADKAFVTVADEPILLIGDVFHHGKETKKNLIENASAEHGVEFGDRW 333
| : : | | | : : : | : : | : : |
DB 211 SDVASIESESEGG---RLVFTVOEKPIITDVVOGSAVSDNLAMSSKKG----- 261
OY 334 LDRSV--DVLTPDNTADVSLYDTGTOYRFDEVVFFITDPKTNLTPTDPDKLPVRELLE 391
262 ---SVISDRLLSQDIQKTTDLIRKEGYLAE--VNYEIKEKENTSSA-----T 304
OY 392 QLLVYNGEAVNLQAVRLASDLATRYFNWVNTETPER-----EQION 437
| : : | | | : : : | : : | : : |
DB 305 LLLTVNEKKLYIKDVRLEGLETIKAK---TLKKELALTERNFLSWTGTGVLREVELER 361
OY 438 DQVFEQSSSRTEPAQVDESTLEP-----VETVELDGLIMDSPIEFSANLQ 489
DB 362 DSIAL---SAYAMNHGYDIQVASEVYTFNKEGIYITFRVKEGKKYKIGKIDFKDLLET 418
OY 490 DKLNLVAAKARHLVMDPDRVLAINHDDGVNRSILGRISDAVASAVARAILDESENE--- 546
DB 419 NEQLLKVTKIDHKNYEQYFSLVWQDD--VKALTFYSDYGAFAEVDL--ETTKNEEDA 475
OY 547 -----VIDLPERPAL-----ANKRTPADY----- 566
DB 476 TIDVTFLLDKKQKVFARRIIVEGNTRTDNVILRELRLADGDLFNGOHLRNSNECLNRLG 535
OY 567 -----OSKKVPLVYFVASDKPRDQIGLGMGSDPTGTRLYTK---FEHNLINR 610
DB 536 YFNQVDITDLPTGKDEDEDLVKKV--QEARTGALTGGYGTSHSGFSGSISEKNLWGK 593
OY 611 DGYQAGAEHLRSEDKKGVKLYATKPLSHPLNDQLRALGYOEYFVGHSTNGFDL----- 664
DB 594 -GYLLIEGFISSKSSSIDLSTFNPRVYDITD-----FGFSNNIYTLDEWMD 639
OY 665 -----STRTLEHISRSIIIONGMNRTYSLRYRLDKLTKQAPETWQDLPVDVNGKP 717
DB 640 FRKKTGDTILFHPRI-----GEYSITFVGRIIDQYRLYLDIPSTAPRSTYLD--QOKN 690
OY 718 SQEALLGAVAHKTVADNLVMPMRGYRQRYSLLEVSSGLVSDANMAIARAGISGVYSEGD 777
DB 691 ISSVVGSGFTDST--DSRERPSKGIHAKLIVERGGGLGNDNFKRIAELOGYSTI-- 746
OY 778 NAYSNRHOMTGGIOAGYIWSDNFNHVPYRLRFPFAGDOSIRGYAHDSLSFISK -GYL 836
DB 747 ---SRSKNHIIHMTRTAAGAAVKNKKRPVPVDRFFIGIDIRGYDITDLAPKDRFGDE 803
OY 837 TGGCVLAVGTAEVNYEFKKDLRLAY--FGDIGNAYDKGFTND---TKIGAGVGRMASP 890
DB 804 IGGRRMAFLNEIYWTQPELGLALVPFYDIGFTDVSQTSNPSKLLQSYGLELRWASP 863
OY 891 VGOVR 895
| : : |
DB 864 MGDLR 868

RESULT 2
US-10-080-505-13
: Sequence 13, Application US/10080505
: Publication No. US20030073166A1
: GENERAL INFORMATION:
: APPLICANT: St. Geme, Joseph W.
: TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
: FILE REFERENCE: A-59941-1/RTT/DEF/DHR
: CURRENT APPLICATION NUMBER: US/10/080.505
: PRIOR FILING DATE: 2002-02-22
: PRIOR APPLICATION NUMBER: US 08/296,791
: PRIOR FILING DATE: 1994-10-25
: PRIOR APPLICATION NUMBER: US 09/839,996
: NUMBER OF SEQ ID NOS: 58
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 13
: LENGTH: 1436
: TYPE: PRT
: ORGANISM: Haemophilus influenzae
```

```
US-10-080-505-13
Query Match 2.9%; Score 139; DB 9; Length 1436;
Best Local Similarity 17.1%; Pred. No. 0.32;
Matches 146; Conservative 138; Mismatches 315; Indels 256; Gaps 36;

OY 29 LAAQNPNANIINHPADHPALINQA--KAGNPVLLPREQIOARLNAAGLNKAPQSOALDVY 87
| : : | | | : : : | : : | : : |
DB 753 LAKLNGVNTLINH--SQFTLSNNAQTQGN---TQLSNNAATVYDANLNG----- 797
OY 88 NFDQSPISRIEQQSPRLQDMSVIEETTPLSLEBLFAQESTEMGINPDYIPRYQGEQP 147
DB 798 -----NVHLTDSAQFSLKNSHFHQIOGDKD 823
OY 148 NSEVVVPTLPEREKGILIKRLAR-----LPNDGVKKVYRLAKKFFQSSQSETSAI 199
DB 824 TTVTLNATWMPSPDATLQNLTLNNTYTLNSAYSASSNNAR-----HRSLETFETPT 878
OY 200 GSSHQKTEPYANIKAALEDITQESAMDINGSIPLRQPLAVARAAGYUDDLSIIRNSI 259
DB 879 SAEH-----RFTLLTVNGKLSG--QGTQFTSSLSFGYKSDKLKLSNDAE 920
OY 260 GEVDYIINDLG--EPVYIDYRAVEVRGEGADK-----AFTTVADEV---PLLGDVPH 309
DB 921 GDTLSVNTGKPEPALBQTLV---ESKDNKPLSDKLKFTLENDHVDAGALRYRLVKN 977
OY 310 GKY-----ETKKNLIENASAEHGVEFGDRWIDRSVDVILPDMTADSLIYDTGTOYR 360
DB 978 GEFRLHNPKEQELRNNDLVRABQAE-----RTLEAKOVEDTALEQ---TSNARVR 1024
OY 361 FDEVFFITDPKTNLTPTDPDKLPVRELLEQLLVNMGAEAVNLQAVRLASNDLIATRY 419
DB 1025 SKRAVFSDTLPQOSQLDV-----LQAEQVEPTAKKKAKKAVSKRAVFSDTLPQOSQ 1078
OY 420 FNMVTEIYVPEREIOIQNDQVSFEQSSSRTEPAPVDESSTLEPYIETVELDGLIMDISP 479
DB 1079 LDVLQAEQVEPTAEKQKKAKKVRSKRAAR-----EFSQD-----TP 1114
OY 480 IEFSSANLIQDLNLVAAKARHLVMDPDRVLAINHDDGVN---SILGRISDAVASAR 536
DB 1115 LLSRLKYLEVLEVINAKQ--QOVKKPEQDEKQKOKDLISRYNSALSATYNSKLS 1173
OY 537 AILPDESENEVIDLPERPALANRKTTPADYQSKVPLVYFVASDKPR--DQIGLGMGSDT 595
DB 1174 --VQDELRLFPVDAQSAVWNT-----INQDRRRDSDAFRAVYQOKT 1213
OY 596 GTRLYTKFEHNLINNDGYQAGAEHLRSEDKKGVKLYATKPLSHPLNDQLR--ATL-----G 649
DB 1214 NLR-----QIGVQKALANGRIGA--VFSHRSRDNTFDEQVKNHATLTMMSG 1257
OY 650 YQOEYFGHSTNGFD---LSTRLEHISRSITQ---NGGMNRTYSLR----- 690
DB 1258 FAQYQMGDLQEGVAVNGTGISAKMAEQQSRKIHRAINYGVAASQFRLQGLQIOPYEYV 1317
OY 691 -----YRLDKLTKQAPPETWOD-----LPVDFVNGKPSQALLAGVA--- 727
DB 1318 NRYFIERENYQSEEVKVTPLSLAFNRVNAIGRVDTFTPTDINISVXPFFVNYVYVSNAN 1377
OY 728 ----VHKTVADNLVMPMRGYRQRYSLLEVSSGLVSDANMAIARAGISGVISGDNAYGSN 783
DB 1378 VQTTVNSTV---LQOPFGRYWK---EVLKAEILHFLDS-----AFISKSQSQ 1421
OY 784 RAHQMTGGIOAGYIW 798
| : : | | |
DB 1422 LGKQONVGVKLGYSW 1436

RESULT 3
US-09-815-242-5635
: Sequence 5635, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
```

```
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR APPLICATION NUMBER: 2001-03-21
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5635
LENGTH: 2368
TYPE: PRF
ORGANISM: Staphylococcus aureus
US-09-815-242-5635
```

Query Match 2.9% Score 137.5; DB 10; Length 2368;  
Best Local Similarity 17.4%; Pred. No. 0.89;  
Matches 202; Conservative 146; Mismatches 423; Indels 387; Gaps 47;

```
QY 2 SKPVLEFANSFPMVALAAYPLMTSQLAQONNPANT-----INHPAHMTALNOAKAGN 56
DB 471 TREVLFPDMS-----LKLSTKYVNVANIDTPKKNIDENEKLTFTASDIVINNO--- 518
QY 57 PVLVLT--PEQIOARLNAAGLNKKPSQA-----L 84
DB 519 PEVTLTADDFSVAYAEVEMKNKDALQOVNSQVNSHYTTASTAEYKCLKQOADNIILEDANHY 578
QY 85 DVVNFDQSPI-----SRIGEOSPPLGLDMSVIEETTPULSELEAQES-----TE 130
DB 579 ETANRASQAIDGLVTKLOAAALIDNOAAIAELDAKAEKVTAQOSKKVTQDEVALVTK 638
QY 131 MGINPNDVPEY-----OG---EOPN-----SEVVVPPTLEPE-KPGLIKRLIYAR---- 171
DB 639 INNDKNNAIAELINKOTTAQGVTEKONGIAVLDDQVITPTVKQAQODIIQAVTTTKOOI 698
QY 172 -----LFDNGVKNVPRLKAKEYOSSQSG-----ETSAIGSSHOKTEPYANIK 213
DB 639 KKSNASLDERKDVANDKICKIEFKAIKIDDAFTNNOVAIKTKALINDINO-TAPATTK 757
QY 214 -AALR-----DITQESAMDINGSTIPRLQGTALVAARAVGYDI--DLS 253
DB 758 AALAEFEEVVOAQIDQAPLNPDTTNEEVAE--AIERINAKVSGVKAIEMTTTQODLE 814
QY 254 IIRN-----SIGEDVVIHIDIG 270
DB 815 RVKNEIEIKIENTITDSTQTKMDAYEVKQAATARTQONATVSNATNEEVAEADDAVEAAQ 874
QY 271 EPPYIDYRAVEVREGADDKAFTTVADEVPLT-----IGDVFHGGYEKKKNLIENA 322
DB 875 KQGLHDIOVNVSKQGVADTK--SKYLDKIMAIQJQAKVPADTVEVNAVYNTRKQEOINS 932
QY 323 SAE-----HGFDGMLDRBSVDVILPNDTADVSLIYDTG-----TOYRFD-- 362
DB 933 NASYTEEKQAAYTTELDTRKQEARFTNLDAANTNSDVTFTAKDNGIAIINOVAATTKKSOAK 992
```

```
QY 363 -EVEFTIDPKT-----NQLTDDPKLPVKKRELLBOLLTVNKGAVYNIOAVRALSNDLIA 416
DB 993 AEIAQKASERTATIEAMNDSTEEQO--AAKDQVQAVVTANADIDN-----AAANTVDVN 1046
QY 417 TRYFNMTVELIYFPE-----REQIONDOVSFQOSSSRKE-----PQVDES 458
DB 1047 AKTTNEATIAAITTPDANVKPPAKQAIADKVOAQETAIIDANNGATTEEKAARKQOVTERK 1106
QY 459 TLEPVIE-----TWELTDGILNDISPIEF-----SASNLIQDKLNLVAKARHLYDM 505
DB 1107 TADPAIDGAHTNAEVEAKAKNEIAKIEAIQPAITTKDKAKAIAIKAEKERTAIQAOTDI 1166
QY 506 PDDRYLAINHDDGVNRSLIGRISDAVSAVARAILPDESENEV-----IDLEPRTALA 557
DB 1167 TAEETIAAN-----ANVDNAVTDQANNNIEAANSQNDVDQAKTTGSEASIDQVPTV 1216
QY 558 NRKTPADYQSKKVPILYFVASDKPROQIGLGMGSDGTGLVTKFEHNLINRQGYQA-- 615
DB 1217 NKKATAVTDAKNNI--TAAITDNGVDTAKDAGKNSIQSTOPATAVYKSNMKN-DVDQAVT 1272
QY 616 -----GAELRLSEDKKGVKLYATKPLSHPLNDQLRATIGVQQGVFGHSTNGFDLS 665
DB 1273 TQNOAIDNTGA--TTEEKMAADVLVAKAEKAYQDILNA-----QTTND-- 1315
QY 666 TRTLEHISRSIIONGWNRTYSLRYRLDKLTQAPPE-----TWODLPVDFVNGKP 717
DB 1316 -----VTQIKDQAVADVQIGTADTTITKDVAKBELATKA 1348
QY 718 -SQBALLAGVAVHKTADVADNLVNPMBRGYRQRISLEVSGSGLVSDANMAIARAGISGVYSPG 776
DB 1349 REQKALTAQPADATTEEKEQANOQ-----VDALETQGNONINEMAQSID 1391
QY 777 DNAYGSNRHQMTGGIOAGYIMSDNFNHPYRLRFAGGDSIGYAVHDSLPSIDMGYL 836
DB 1392 DVTMAKNMAIQALDIPDIASTVDKTN-----ARAEMLTEMOKKTIELIINNMTNEEG-- 1444
QY 837 TGGCVLAVGTAEYNEFMKDLRLAVFGDIGNAYDKGFTNDTKIGAGVGVWASPVQGVRY 896
DB 1445 -----NDI-----GTVRAAYEEGLNN-----INAAITTGQVTT 1472
QY 897 DVATGVKE-----EGNPVK 910
DB 1473 AKDTAVQKVQOOLHANPYK 1490
```

RESULT 4  
US-09-815-242-12389  
Sequence 12389, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Carr, Grant J.  
APPLICANT: Trawick, John D.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR APPLICATION NUMBER: 2001-03-21  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625

```
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 12389
;; LENGTH: 2368
;; TYPE: PR
;; ORGANISM: Staphylococcus aureus
US-09-815-242-12389

Query Match
Best Local Similarity 17.4%; Pred. No. 0.89; DB 10; Length 2368;
Matches 202; Conservative 146; Mismatches 423; Indels 387; Gaps 47;

2 SKPLFANRSFMPVALAAYLPLMTSQLAQQNNPANI-----INHPADHAIINQAGN 56
Db 471 TKEVLFPPKS-----LALSYKVNANIDTPKIDFNKLTFRASDIYINNAQ--- 518
Qy 57 PPVLT--PEQIOARLNAGLNAPQSA-----L 84
Db 519 PEVLTADPFSAVAVEMNKDALQQVNSQVNSHYTTASIAEYKMLKQADNINLEDANHV 578
Qy 85 DVNFDQSP-----SRIGOSPPLGLDMSVIEETTPLSLEELFAOES-----TE 130
Db 579 ETANRASQAIDGLVTKLQALALIDNQAIAELDAKAEKVTAQAQSKVTQDEVALATK 638
Qy 131 MGINPNDYIPEY-----QG---EQPN-----SEVVPPLEPE-KPGLIKRLVAR--- 171
Db 639 INDKNNIAIEINQGTQAGTTEKNGIAVLDOVITPVPKAKODIQAIVTRKQOI 698
Qy 172 -----LFDGYNKPKLAKAFYQSSQ-----FTSAGSSHQTEERYANIK 213
Db 699 KKSNAIODEKDVANDKIGKIEKAIKIDATTAQVNAQVEAIRTKAINDIQ--TAATYAK 757
Qy 214 -AALE-----DITQESAMDLNGSIPRLQOTALVAARAVGYDI--DLS 253
Db 758 AAALPEPEVVOAQIDQAPLNDPTNEEVAE--AIERIMAKVSGVAIAEITTAQDLE 814
Qy 254 IIRN-----SIGEVDVITIHDLG 270
Db 815 RYKNEISKIENITDSTQTKMDAYNEVQQAATARKTONATVSNATNEEVAEADAVEAAQ 874
Qy 271 EPVYIDYAVVEGRGADKATTYADEVPL-----IGDVHHGKYEKKULIENA 932
Db 875 KQGLHDIOVAKSKEVADTK--SKVLDKINAIQTAQAKVPAPADTEVENAYNTRKQEIQNS 932
Qy 323 SAE-----HGYPDGMILDRSVIILPDMTADVSLIYDTG-----TOYRFD-- 362
Db 933 NASTTEEQAAVTELDTKKQAEARTNLDAANTNSDVTAKDNIAAINQQAATTKKSDAK 992
Qy 363 -EVEFTDPTK-----NQLTTPDKLPYKRELEQLLTVNKGAVNIQAVALSNDLIA 416
Db 993 AEIAQKASERKTAIRAMNDSTTEEQ--AAKDKVQAVVTAANADIDN--AAATVDVN 1046
Qy 417 TRYFMVMTETIYEP-----REQIONDOVSFEQSSSSTE-----PAQVDS 458
Db 1047 AKTTNEATIAITTPANAKPPTAKQALADVVQAQETAIIDNNKATTEEKAAKQOVTEKT 1106
Qy 459 TLEPIE-----TVELTDGLMDISPIEF-----SASNLIDQKLNVAARHLYDM 505
Db 1107 TADTAIDGHTNAEVEAAKNMAIEAIOPTATTKDNKAKQALANRANRKAIRTAIAQTOI 1166
Qy 506 PDDRVLAINHDDGVNRSLIGRISDAVSAVARAILPDESENEY-----IDLPERTALA 557
Db 1167 TAEETIAAN-----ANDNAVTQANNNIEANSONDDQAOKTTGEASIDVTPV 1216
Qy 558 NRKTPADYVQSKVPLVYFVASDKPRDQIGLQWGSQDTRLVTFEHLNIRDGOA-- 615
Db 1217 NKKATAVTDAKNNI--TAATIDNGVDTRAKDAGKNSIGSTQATAVKSNAKN-DVDQAVT 1272
```

```
Qy 616 -----GAELRLSDKKGVKLYATKPLSHPLNDQLRATLGLQGEVFGHSTNGFDLS 665
Db 1273 TONQALDWTGA-----TTEEKNAAKDLVLKAKEKAVQDILNA-----QTTND----- 1315
Qy 666 TRLEHEISRTSIQNGWNRRTYSLRYLDDKLTKQAPPE-----TWQDLPVDVNGKP 717
Db 1316 -----VTOIKDQAVAVQGITADPTTKDKVAKDELATKA 1348
Qy 718 -SOEALLGAVVHKTVDNLVNPMBRGYRQRYSLFVSGSSGLVSDANMAIARAGISGVSG 776
Db 1349 REQKALINQIADATTEEEQANQ-----VDALLTQGNQNIENAGSID 1391
Qy 777 DNAGVSNRAHQWTGGIQAQYIWSDNFNHVPYRLREFAGDOSIRGVAHDSLSPIQDGL 836
Db 1392 DVNTAKDNAIQALIDPTQASTDVKTN-----ARAEILTEMQNKITELINNFTNEKG-- 1444
Qy 837 TGGQVLAVGTAETNTEPKDLRLAVFGDIGNAYDKGFNTDRTIGAGVGRASPPYGOVRV 896
Db 1445 -----NDI-----GPVRAAYEEGLNN-----INAAFTTGQDVTT 1472
Qy 897 DVATGVKE-----EGNPIK 910
Db 1473 AKDTAVQKQQLHANPVK 1490

RESULT 5
US-09-815-242-11571
Sequence 11571, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Twawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes In
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 11571
LENGTH: 2890
TYPE: PR
ORGANISM: Helicobacter pylori
US-09-815-242-11571

Query Match
Best Local Similarity 19.4%; Pred. No. 1.5; DB 10; Length 2890;
Matches 198; Conservative 145; Mismatches 361; Indels 318; Gaps 52;

Qy 55 GNPVLTPEQIOARLNAGLNAPQSAQALDVVNPFDQSPISRIQOSPPLGLDMSVIEE 114
Db 1675 GAPEITVRNEK-----RMLQEAADVLFQDGRSTNAYKANKKRP----- 1713
```



```
Db 1765 -----RLAAKHLVEQALNOAIDQINHA-----DRTAOVNODSINAQINISIKRPAT 1810
Qy 532 SAVARAI-----LPDESENEVIDLPE-----RTALANRKPADYV 566
Db 1811 TVKATALQOIONATNKNLIKANNENATDEQNIAlAOVEKELIKAKQOIASATYNADV- 1869
Qy 567 OSKRVPLVYVASDKPRDQIGLWGSPTGTRLYTKFEHNLINRDGYQAGALRLSEDK 626
Db 1870 -----AYLHDEKNE-----IREIE-PVINRKASAREQLTTLFNDKK 1905
Qy 627 GVKLYATKPLSHPLNDOLRATLGYOEFGHSTNGFDLSTLTLEHISRTIIONGWNRT 686
Db 1906 -----QAIEANIQATVEERNSTLALQLONIYDIAIQIDODRS-----NAQVXT 1949
Qy 687 YSLRYRLDKLKTQAPPEWODLPVDFVNGKPSQEQALLAGVANHKTVDNLVNP--MRGY 743
Db 1950 ASLNL-----QTIHDLVHPIK-KPDAE-----KTINDLARVTALVQNY 1988
Qy 744 RQRYSLLEVSSGLVSDANMAIARAGISGV 772
Db 1989 RK-----VSNRNKADALKAITLAL 2006

RESULT 7
US-09-815-242-12913
; Sequence 12913 Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 12913
; LENGTH: 2186
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-12913

Query Match 2.9%; Score 135; DB 10; Length 2186;
Best Local Similarity 19.0%; Pred. No. 1.2;
Matches 154; Conservative 110; Mismatches 301; Indels 244; Gaps 35;
```

```
Db 1421 IQPETHVPAAREKTIN-----QKANEIRAKINODKEATAEROVALDKINEFVNOAMT 1473
Qy 130 EMGIN-PNDYIPEYOGEOPNSEVVPPTLEPEKPGILKRLYARLFENDGVNKPRLKAFY 188
Db 1474 DITNRTNOQVDDTISQALDSIALY-----TPPHIVAAAR---DAVKQ--QYEAKR 1521
Qy 139 OSSGGETSAGSSHQKTEPEYANITKALEDITQESAMLDNGSIPRLROTALVAARAYY 248
Db 1522 EIEQAEHATDEEKQVALNOLANNEKRALQNIQAI-----NNQVKRETEGTATLKGV--- 1575
Qy 249 DIDLSITNSIGEVOYIITHDGEPEYIDYRAVEBGAGDKAFYTV--ADVEPLIGGV 306
Db 1576 -----QPHIVIKP--EAQQAIKASENOVESIKDIPPHATVDELDEANQLISPT 1621
Qy 307 FHHGKETKKNLIENASAEHGYFGR-WIDRSVDYLPD---NTADVSLIYDTGTQYRPD 362
Db 1622 LKQAOQE-----IENTNQDAATDVNRQTKAIEDQKPRVRRKRALDSENNKN--- 1672
Qy 363 EYVFTIDPKTNQLTTDPDKLPEVKRELLBQLITVMNGEAYNIQAVRALSNDLIATRYFM 422
Db 1673 -----QLDAIRNTLDTTDQDERDVAIDTLNKI-----VNTIKNDIAQ-----NK 1710
Qy 423 VNTETVPEREQIONDOY-----SPEOSSSRTEP--AQVESTLEPEYETVELD 471
Db 1711 TNAEYDRTETDGNDAIKYILPVQYKPPARQSVGYKAEQNALIIOQSLSTEE----- 1764
Qy 472 GILMDISPIESASNLIDCKNLVAKAKHLYMDPDRYLAINHDGVNRSILGRISDAV 531
Db 1765 -----RLAAKHLVEQALNOAIDQINHA-----DKTAQVNOQDSINAQINISIKRPAT 1810
Qy 532 SAVARAI-----LPDESENEVIDLPE-----RTALANRKPADYV 566
Db 1811 TVKATALQOIONATNKNLIKANNENATDEQNIAlAOVEKELIKAKQOIASATYNADV- 1869
Qy 567 OSKRVPLVYVASDKPRDQIGLWGSPTGTRLYTKFEHNLINRDGYQAGALRLSEDK 626
Db 1870 -----AYLHDEKNE-----IREIE-PVINRKASAREQLTTLFNDKK 1905
Qy 627 GVKLYATKPLSHPLNDOLRATLGYOEFGHSTNGFDLSTLTLEHISRTIIONGWNRT 686
Db 1906 -----QAIEANIQATVEERNSTLALQLONIYDIAIQIDODRS-----NAQVXT 1949
Qy 687 YSLRYRLDKLKTQAPPEWODLPVDFVNGKPSQEQALLAGVANHKTVDNLVNP--MRGY 743
Db 1950 ASLNL-----QTIHDLVHPIK-KPDAE-----KTINDLARVTALVQNY 1988
Qy 744 RQRYSLLEVSSGLVSDANMAIARAGISGV 772
Db 1989 RK-----VSNRNKADALKAITLAL 2006

RESULT 8
US-09-765-272-118
; Sequence 118 Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 118:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1040 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 118:  
US-09-765-272-118

Query Match 2.8%; Score 134; DB 10; Length 1040;  
Best Local Similarity 19.4%; Pred. No. 0.44;  
Matches 190; Conservative 124; Mismatches 318; Indels 348; Gaps 53;

QY 24 MTSQALAOQNNPANI--INHP-----AHDTAINQAKAGNPVLLTPEOIOARLNKA 73  
DB 52 LTSQLSAANSQSLSIGVGHLEPPLKIEGYIGYIKTKKQNTLSRT---VDGKYSQ 108  
QY 74 GLNAKQSQALDVNFD-----QSPISRIGESPPGLSDMSVIEETPLSLEELFAOE 127  
DB 109 RDSQNSTSTSVSHASADEMMNOGKVSLOGEASGDDG-----SEKSIADNLSND 163  
QY 128 STEMGINPDYIPBEOGEOPNSEVVPPLLEPKPLIKRLVRLNDGVNKPRLKAF 187  
DB 164 SASQVBN--PDHGES-----VVRPTV-PEQGNPVA----- 194  
QY 188 YOSOSGETSAIGSSHOKTEPYANIKAALEDI--TQESADLNGSIPRLQOTALVARAVG 246  
DB 195 -TTVQSAEEVALATNDRE-----YKLPLETKGTQEPGHE----- 229  
QY 247 YYDIDLSTIRNSIGEVNIIHDLGEPVYIDYRAVEVRG--EGADDKAFTVADEVPLLIGD 305  
DB 230 -----GEA-AVREDL--PVYT--KPLETKGTQEPGHEGAADVREPEAYTER 271  
QY 306 VF-----HHGXYETKKNLE-----NASEHGTFDGRMLDRSYDVLIPONTAD 348  
DB 272 LATKGTQEPGHEGKATVREETLEYEPVATKGTQEPH--EG--ERYVEELP----- 320  
QY 349 VSLIDTGTQYRFDEVFETIDPKTNQLTTPDPKLPVKKRELEQLLTVMNGEAYNLQAVR 408  
DB 321 -ALEVTTNRRTIQNI-----PYTTEIQDPTLLKNRKRIEQ-----GQA----- 360  
QY 409 ALSNDLIATRYN---MVNTEIVPERBOIQNDQVSFEQSSSRTEPAQVDE---STL 460  
DB 361 -----TFRTIQYEDYIVNGNV-----ETKEVSRPEVAPVNEVAVKVL 399  
QY 461 EPVITVELTQIIMDISIEFSAS-----NLIQDKLNLYAAKARLLYMPDDRVLAINH 515  
DB 400 VKVKTVEIT-----NLKVENKKSTSYNLIDTTSAYSAKQ-----VFH 442  
QY 516 DDGVNRSILGISDAVSAVARAILPDESENEYIDLPERFALANRKTPADVYQSKVPLV 575  
DB 443 GD-----KLKVEYDIENPAKEQVI-----SGLDYIPIYTKTHL 476  
QY 576 FVASDKPRDQIGIGMGSDGTGLVY--KFEHNILNRDGOYAGABRLSEDKGKYLVA 632  
DB 477 -----TYNLGNNENETETSTQDFQLEYKIE-----IKDIDSELVLG 514  
QY 633 TKPLSHPLNDQIRATLGLGOQEVFGHSTNGFDLSTRLEHISRSIIONGGMNRTYSLRR 692

DB 515 KE-----NDIRYRYSLS-----EAPDTAK-----YFVKV 541  
QY 693 LDKLQAPPETWQDPEVDFVNGKPSQEALLAGVAHKTVADMLVNMGRYRORYSLEVG 752  
DB 542 SDRFEKEM-----LPVKSIT--ENTDGTYKVTAVADQVLEE--TDGYKDDYFTTVA 589  
QY 753 SS-----GLVSDANMAIA--RAGISGYYSFGDN-----AYGSRAHQMTGGIQAGTWS 799  
DB 590 KSKAEOPGVYTSFKOLVATAMQSNLSGVYTLASDMTADSEVSLGDKQTSYLNGAFTGLSIG 649  
QY 800 DNF--NHVPYRLR-----EFAGG--DOSIRGYAHD-----LSPISDKY 835  
DB 650 DGKSTAIYDLKPLDPTLNGATVRRIDITVSADESKENVAALAKAANSINNVAVEGK 709  
QY 836 LTGGOVL-----AVGTAEVYEFMKDLRLAVFGDIGNAVYDKGTNDTKIGAGVG--VRW 887  
DB 710 ISGAKSVAGLVASATMTVIENSSFTGKL-----IANHDSN-KNDT--GGIVGNITGN 759  
QY 888 ASPVGOVRVDVATGVKEEGN 907  
DB 760 SSRVKNRVVDALISTNARNN 779

RESULT 9  
US-10-160-758-15  
Sequence 15, Application US/10160758  
Publication No. US20030036076A1  
GENERAL INFORMATION:  
APPLICANT: EXELIXIS, INC.  
TITLE OF INVENTION: CADS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE  
FILE REFERENCE: EX02-089C  
CURRENT APPLICATION NUMBER: US/10/160,758  
CURRENT FILING DATE: 2002-06-03  
PRIOR APPLICATION NUMBER: US 60/296,076  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: US 60/328,605  
PRIOR FILING DATE: 2001-10-10  
PRIOR APPLICATION NUMBER: US 60/357,253  
PRIOR FILING DATE: 2002-02-15  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 15  
LENGTH: 4349  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-160-758-15

Query Match 2.8%; Score 133.5; DB 9; Length 4349;  
Best Local Similarity 17.4%; Pred. No. 4.4;  
Matches 199; Conservative 162; Mismatches 388; Indels 395; Gaps 50;

QY 52 AKAGNPVLLTPEQ--IQARLNAGLNKQSQALDVNFDQSPISRIGESPP--LGID 108  
DB 2145 ARDGGTPELQSEEEVLTVYNNKSNPLFQSPYKVRVPENTTLVYPIHMQARSPGRLI 2204  
QY 109 MSYIEE-----TTPLSLE---ELFAQSTEMGINPDYIPBEOGEOP 147  
DB 2205 YNIVEEPEPLMLFTTDKTKGVLTVTGPLDYESKTKHFTVATATQALG-----SESEA 2256  
QY 148 NSEVVV-----PPTLE-----PEKPLIKRLVA-----RLENDGV 177  
DB 2257 TVEVLVEDVNDNPPFQSQVLYTTSISEGLPAQTPVLIQLLASDQSGNRDVSQIIVEDGS 2316  
QY 178 NKVPRLKAFYQ--SSQSGETSAIGSSHOKTEPYANIKAALED-----IQESAM-----D 226  
DB 2317 D-----VSKFFQINGSTGEISTVQELDYEQQHFNKAVRAMDKGDPPLVGETLVVNVSD 2371  
QY 227 LNSGIRPLQOTALVARAVGYUIDLSIIRNSIGEVNIIHDLGEPVYIDYRAVEVRG-- 285  
DB 2372 INDNPEPFQPOYEANVS-----ELATGCHLVKQALIDPDRDSRLLEYLLISGNQR 2425  
QY 286 -----GADDKAFTVADEVPLLIGDV-----FH 308

```
Db 2426 HEFTNSSGII SMFLCKKHLSDSSYNI RVGASDGVFRAT---VPVYINTNANKYSPEEQ 2482
QY 309 HGKYEYTKKNLEIASAEHGYEDGRLMDRSVYVILPDNTADVSLIYDTGTQYRDEVVEFT 368
Db 2483 OHLYEAE--LAENAMVGTKVIDLAIIDKDSG---PYGTIDITTIINKLASEK-----FS 2530
QY 369 IDPKTNOLTTDPKLPVKRELLEOLLVNMGEAYNLQAVRALSND-----413
Db 2531 INP-NGOIAI-----LQRLDRENSSTE--RVIAIKVMADGGGRVAFCTVILIT 2576
QY 414 -----LATRYENMNT-----ELVPPERQIONDOVSESSSSSRPEPAVD 456
Db 2577 DENNPPQFASETTVTSIQSVNSKDSPIYQVLAVDAGQADVYI---SVNPEDLVKD 2632
QY 457 ESTLEPIETVELTDGIL-MDISPIEFA-----SNLIDKYLVAAKAR-----500
Db 2633 VIEINPYTGAVKVDLSVGLENQFLDEFIKAQDGGPRHNMNSLVRLQVVPKKSYPKFS 2692
QY 501 ---HLVMPDRDLAINHDDGVNSILGRISDAVSAVARAILPDESENEVIDLEPRTALA 557
Db 2693 EPLTFSAPED--LPEGSEIGIVAKAAQ--DPVYISLVKGTTPESNKKDGVFSLDPDTCVI 2749
QY 558 NRKTPAD-----VYQ-----SKK 570
Db 2750 KYRKPMDHESTKLYQIDVMAHCLONTDVSLVSVNIQGVNDNRKPVFEADPYKAVLTEN 2809
QY 571 VPL-----YFVVASDKPRDQIGLGMSDGTGRLVTKEHNLINRD-----611
Db 2810 MPVGTSVIQTALIDKIDTGRGOVSVRLSADPGSNV-----HELPAIDSESGMITTLOELD 2864
QY 612 -----GYOAGAEIRLS-----EDKKGVKL 630
Db 2865 CENCOTYHFHVAVADHCQTTQLSSQALVOVSIITDENNAPFASEEVYRGSVVENSEPEGL 2924
QY 631 YAT-KPLSHPLDQRLATLGYOE---VFGSTNGFD--LSTRLEHISRSIIQNG 681
Db 2925 VATLTGLDADISEQNRQVTCYITGDPGLGQFISQVGDENMISRSKTLDRHTAK-----2979
QY 682 GMRRTSLRLRLDKLTQAPRPETMODLPVDFVNGKPSQEBALLAGVAVHKTYADNLVNMIR 741
Db 2980 ----HLRTYASDGKQA--SVYVEIFVLVNDNSPQCSQLLYTKKHEDV-----3024
QY 742 GYRQYSLEVSSGLVSDAMMAIARAGISGVYSPGDNAVSGNSRAHOMTGGIOACIYS--799
Db 3025 -FPGHFILKXASANDLDDTNAQI-----TYSL-----HGBG-AHEKRLDPHTELTTLT 3071
QY 800 -----DNFNHVPYRLRFFAGDQSTIRGYAHDLSPISDKGYLTGGQVLAVGTAEVNYE 852
Db 3072 ALDRERKDVFNLV---AKATDGGGRSCQADITLHVEDVNDNA-----PRF 3113
QY 853 FMKDLRLAVGCD--IGNAVYKGFINDTKIAGVGVNRASP---VGOVRYVATGVKEEGN 907
Db 3114 FPHCAVAVFNDTTVKTTPAVAVFARPDQGANOVVSLPDSAGHPSIDATGTVIRLEK 3173
QY 908 PRL 911
Db 3174 PLQV 3177

RESULT 10
US-09-815-242-12610
; Sequence 12610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
```

```
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12610
; LENGTH: 5795
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12610

Query Match 2.8%; Score 133; DB 10; Length 5795;
Best Local Similarity 16.7%; Pred. No. 7.5;
Matches 165; Conservative 144; Mismatches 348; Indels 332; Gaps 45;

QY 51 QAKAGNPVLLTPEQIQAR-----LNAAGLNAKQSQALDVNFDQSPISRT---GEQ 101
Db 792 RVPKPTPTITTTAEQLRGTALOKVPNVTSIGIPLDPSALVYL-----APTQNTNGSE 845
QY 102 SPLLGLDMSVIEETPRLPLELEFAQESTEMGNPNDY---IPEYGEQ-----146
Db 846 ADQIPSGTIIATGPDGVNHTIT-----IQPDIVYFIPPV-GKQIRAVYVYNNKYVA 897
QY 147 ---PNSEVVVPPTLEPERGLIKRLYARLFNDGVKNVRLAKFYQSSQSGEISAIGSH 203
Db 898 SNMSNAVPIILDDIPPT-----INNPVGINAKYRGDEVNFTMGVSDBH 941
QY 204 Q-KTEPANIKAAL-LEDTIQESAMDLSIPRLKQTLALVAAVAGYVDIDLSITRNSIG 260
Db 942 SGLSTYITTLTPSGWTSNLTFSKSD--NKNGLSLATIGRVASMNA---FNSDITFRVSATD 994
QY 261 EVDYIHDIGEPVYIDYRAVVEVRGEGADKAFYVADEVPLLIGDFHHGKETKKNLIE 320
Db 995 NVNMTTND-SQSKHVSINVKIS-----EDAHPIVLGN-----PEKVVVV 1033
QY 321 NASAE-----HGYPDGRWLDKRSVDVILPDNTADVSLIYDTGTQYRF 361
Db 1034 NPTAVSNDEKORITTAFMKNKQNIKRGYL-----ASTDPVYVNDNGNVTLHVRDSSSTTL 1087
QY 362 DEVEFFETIDP--KTNOITTPDKLPVKRELLEOLLVNMGEAYNLQAVRALSNDLIARV 419
Db 1088 DATVMTTEPYVVKSEYOTANAKTAT-----VTAKGGSFNGIDK-----OY 1130
QY 420 FNMVTEIVPERQIONDOVSESSSSSRTEP-AQVDESTLEPIETVELTDGILMDS 478
Db 1131 FTLSNGQAI-----PSGTFTNITSDRITPTAQ-----1157
QY 479 PIESASNLIDOKILNVAAKRHLIDMPD-----RVLAINHDDGVNR---SILGRISDA 530
Db 1158 --EVSQNNAGTOLYHIVASNAHY--KDEDFYISLKIYDVQKQPEGQRYRSTYDLTTDE 1214
QY 531 VSAVARAILPDESENEVIDIPE-RTALANKRTPADVY-----566
Db 1215 ISKYKQAFI--NARNDYITLAEGLISVTNTPNGANVSTIYVINKGRITKFSASNLAMN 1272
QY 567 -----QSKKVPVLYFVVASDKPRDQIGLGMSDGTGRLVTKEHNLINRDGYOAGAE 618
Db 1273 FLRWVNPQDVTYVWTKIANRPTDG--GLSWSD-----HMSLIYRYDATLTGTQ 1321
```









PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 12967  
 LENGTH: 2478  
 TYPE: PRF  
 ORGANISM: *Staphylococcus aureus*  
 US-09-815-242-12967

Query Match 2.8%; Score 132; DB 10; Length 2478;  
 Best Local Similarity 17.8%; Pred. No. 2.4;  
 Matches 218; Conservative 161; Mismatches 436; Indels 412; Gaps 52;

QY 2 SKPLFANRSEFMPVALAAYLPLMTSQAIAOONPANI-----INHVPADTAIQAQAGN 56  
 DB 475 TKREVLFPDKS-----LKLSTYKVVANIDPFPKNIDFNEKLYFRASDVVINAQ--- 522  
 QY 57 PPVLLT-----PEQIQARLNAAGLNA-----KPSQALDVVAFD--- 90  
 DB 523 PEVTLFADPSVAVENMKDALQOVNSQVDNSHYTTASIAEYNLKKQOADTILMEDANHV 582  
 QY 91 -----DOSP-----SRIGQSPPLGLDMKSVIEETTPLSELEFAQES-----TE 130  
 DB 583 KTAANRASQADIDGLVTKLQALALIDNQAIALDELTKAQEKYTAAGQSKKYVODEVALVTK 642  
 QY 131 MGINPNDYIEY-----QG-----EQPN-----SEVVPPPLEPE-KPGLIKRLYAR--- 171  
 DB 643 INNDKNAIAIEINKQTTAQCVTEKDKGIAVLBEDDVITPVPKPAKODIIOAVYTRKQOI 702  
 QY 172 -----LFNDGVNVPRLKAKFYOSSG-----ETSAISSHQKTEPYANIK 213  
 DB 703 KKSNAISQDEKDVANDKIGIEIKAKIDIAATTAAGVEALIKTAINDINOTTPATTAKA 762  
 QY 214 AALE-----DITQESAMDLNGSIPRLROTAIVARAVGYDI--DLSTI 254  
 DB 763 AALEEFDEVQAQIDQAPLNPDTTNEVAE---AIERINAKVAGVAIEATTQAOLER 819  
 QY 255 IRN-----SIGEDVYIHDIGE 271  
 DB 820 VKNEELSKIENTIDSTQTKMDAYNEVQAATARKAQAATVSNATNEVAEADAVDAQK 879  
 QY 272 PVIIDYRVRGEGADDAKFTTVADEVPL-----IGDVEHHGKYETKKNLINAS 323  
 DB 880 QGHLDIQVAVSKQEVADTK--SKVLCKINAIQQAQKAPRADTVEVNAIYTRKEIOTNSN 937  
 QY 324 AE-----HGFQGRWLDRSVDVILPONTADVSLIYD-----TGQYRFD-- 362  
 DB 938 ASYTEERQAAYTELDTRKKQEARTNLDAANTNSDVTAKONSIAAINQQAATTRKSDAKA 997  
 QY 363 EYVEFTIDPRT-----NQLTDDPKLPVKRELLEQULLTVMMGEAVNIOANRALSNDLIAT 417  
 DB 998 ELQAKSEKRTAIEAMNDSTTEEQ--AAKDQVQAVVTANADIDN---AAANNDVNA 1051  
 QY 418 RYENMVNTEIVPE-----REOIONDOVSEFOSSSSSTE-----PAQVDEST 459  
 DB 1052 KTTNEATTAIATPDANVAKPAQAQIADKVOAQETALIDGNNGSTTEEKAQKQOQTEKTT 1111  
 QY 460 LEVPIE-----TVELTGILMDISPIE-----SASULIDCKLNLVAAKARHLYDMP 506  
 DB 1112 ADAIDAHAHTNAEVAEAKKAIAKIEAIOPATTTTKDNAKEAIAITKANERKTAIAOTDIT 1171  
 QY 507 DDRVLAINHDDGVNRSILGISDAVSVAARAILPDESENEV-----ID----- 549  
 DB 1172 AEETIAAAND-----VDNAVTOANSNITEAANSQNDVDAKTTGGENSIDQVTPTVN 1221  
 QY 550 -----LPERTALANRK-----TPADVQSKKVPVLYEVSADKPRDQOIGLIGWSDTGTGR 598  
 DB 1222 KATATANEITAILNKKLQEIQATPPADTDEKQAA---DAEANTENKKAQALISAATATNA 1277  
 QY 599 LVTKFHNILNDRGYQAGLELRLSE--DKKGVKLYATKPLSHPLNDOLRAATLGYQOEEVGH 657  
 DB 1278 QY-----DEAKANEMAIINAVTPKVVKKQAQAKDEI---DQLOAT---QTNVINN 1320

QY 658 STNGF-----DLSTRILE--HEISRSITONG-----GNNRTYSLRYRLDKLTQ 699  
 DB 1321 DQNAATTEKEKAIIQIALAVYDAKKNITATATDDNGVDQAKQAKNSIQSTQ--PATAVKSN 1379  
 QY 700 APPE-----TWQDLPVDFVKNRKPSEALLAGVAVHKT-----VADN 735  
 DB 1380 AKNVDQAVTTQNOAIDTPTTATTEEKNAARDLVYLAKEKAYODILNAQTNDVTQIQQ 1439  
 QY 736 LVNPMRGYRQRYSL-EVGSGLVS-----DANMAIARA 767  
 DB 1440 AVADIQGTAITATTIKDVAKDELATKANQKALIAQTADATTEKEQANOQVDAQULQGNQ 1499  
 QY 768 GISGVYSFGDNAYSNNRAHQMTGGIQAQYIWSDNFNVHVPRLRFPAGGDSIRGYADSL 827  
 DB 1500 NIENAQSIDVNTAKDNAIQALIDPLQASTDVKTN-----ARAEILLTEMQNKITETILNNE 1554  
 QY 828 SPIDSKGYLTGGVYLVAGTAEYNEFMKDLAVFQDIGNAYDKGFTNDTKIAGAVGRW 887  
 DB 1555 TTNEKG-----NDI-----GPVRAAYEEGLNN-----INA 1580  
 QY 888 ASPVGOVRVDVATGVKE-----EGNPRIK 910  
 DB 1581 ATTGGDYTTAKDTAVQKVOQLHANPVK 1607

RESULT 16  
 US-09-815-242-5703  
 Sequence 5703, Application US/09815242  
 Patent No. US20020061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Karl L.  
 APPLICANT: zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OR INVENTION: Identification of Essential Genes in  
 FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: Prokaryotes  
 PRIOR APPLICATION NUMBER: US/09/815,242  
 PRIOR FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 5703  
 LENGTH: 2025  
 TYPE: PRF  
 ORGANISM: *Staphylococcus aureus*  
 FEATURE:  
 NAME/KEY: VARIANT  
 LOCATION: (1)....(2025)  
 OTHER INFORMATION: Xaa = Any Amino Acid  
 US-09-815-242-5703

Query Match 2.8%; Score 131.5; DB 10; Length 2025;  
 Best Local Similarity 19.6%; Pred. No. 1.9;  
 Matches 187; Conservative 130; Mismatches 337; Indels 301; Gaps 49;

```

QY 48 AINAK---ACNPPVLTPEIQARL-NAAGLNKPOSALD--VNFDDSPISRIGEQ 101
   ||||| : : : : : ||||| : : : : :
Db 315 AVNDKNNLHGDKLADOKRATETLNNLSLNT-PQRQALENOINNAATGEVAQKLE 373
QY 102 SPPGLMSVIEETPLSLELFAOESTEMGINPDYIPE-----YGEQPNSEVVP 154
   : : : : : : : : : : : : : : : : : : : : :
Db 374 AQALNOMEALRN---SID---QOQTEAG---SKFINEKPOKDAYQAQVAKKLDIN 423
QY 155 PLEPERKGLIKRLYARLFNDGVNKP---RLKAKFYOSSGSEI-----SAIGSSHOK 205
   ||||| : : : : : ||||| : : : : :
Db 424 QTSNP-----TLDKAQVEQLQGVNAQAKNLDGDKLADKQHAVTDLNQLNSLNNPQRA 479
QY 206 TEPRANIKALEITQE---SAMD-----LNGSIPRLQOTALVARAVGYIDISIR 256
   ||||| : : : : : ||||| : : : : :
Db 480 LESQINNAATREVAQKLAEQALDQAMQALRNSIQOQOTE-SGSKFINEKPOKDAYQ 538
QY 257 NSIGEVNVIITHDGEPPYIDYRAVEVEGEGADKAFITVADEVPLLIGDFVHGKYE--- 313
   : : : : : : : : : : : : : : : : : : : : :
Db 539 AAQVQAKDLINQGNP-LDKSOVEOL-----TQAVTTAKNML-----HGDQKLAR 583
QY 314 -----TKKULIENASAEHGFDFGRMLDRSVYILPD--NTA-----DVSILYDTGTQY-- 359
   ||||| : : : : : ||||| : : : : :
Db 584 DQQAQVTTVALPN-----LNHAQQAALTDALINAAPTREVAQHVOGTATELDH 631
QY 360 -----REDEVVFTTIDPKTNOLTTDPDKLPVKRELLEQLTV-----NMGEAVNIQ 405
   : : : : : : : : : : : : : : : : : : : : :
Db 632 AMETLKAKVQVNTDKAQPNYTEASTD-----KKEAVDQALQAAESTIDPTGNSNANKD 685
QY 406 AV-BALSNDLIATRYFMNMTETIYFPEREQION--DOVSEGOSSSRTPEAOVDEST-LE 461
   ||||| : : : : : ||||| : : : : :
Db 686 AVEQALITK--LOEKVVELNGENERVAEKAQAKOTIDOLAHLMNQDITAKONIDQATKIQ 743
QY 462 PVIEVELTDLIMDISPIEFSASNLIDQKLNLYAKARHLVMDPDRVLAINHDDGVNR 521
   ||||| : : : : : ||||| : : : : :
Db 744 PIAELVD-----QATOLNSMDQLQQAVERNANVEQ 774
QY 522 ST-LGRISDAVAVARAILPDESEN-----EYIDLPERPALA 557
   : : : : : : : : : : : : : : : : : : : : :
Db 775 TYDVYQADSDKONAKYQALN--EAENVLKQNSKQOVDQALQNTLANKQALNGDERVALA 832
QY 558 NKRTPADYVQSKVPLVYFVASDKPRDGOIGLWGSPTGRVLTKEEHLINRDCYQAGA 617
   ||||| : : : : : ||||| : : : : :
Db 833 KTINGKHDIDQLNL-----NNAQDDGFEKG-- 856
QY 618 ELRLSDKKGVLKATKPLSHPLN-----DOLRATLGYOQVEFGH-- 657
   ||||| : : : : : ||||| : : : : :
Db 857 ---RIDQ-----SHDLNQIQOIVDEKALNRAMDOL-----SOEISGNEGRT 895
QY 658 ---STNGFDSLSTRTLE-----HEISRSIIIONGGMNRTYSLRYRLDKLQAPPETMODLP 709
   ||||| : : : : : ||||| : : : : :
Db 896 KGSSTNVNADTQVQYVDEAVDKAQLDKSTGQNLTAEGVYIKLNDVATAKKA----- 949
QY 710 VDFVNGKP-----SOBALLAGVAVHKTVDNLNVPKRGYRORYSLEVGSSGLVSDANMAT 764
   ||||| : : : : : ||||| : : : : :
Db 950 ---LNGEERLNNRKSSEALOR-----LDQLTHLNNAPQALQIQIINNAETLNKKSRAI 998
QY 765 ARAGISGVYSEFGDNAYGSNRHOMTGGIQAGYIWSDNFNHVPRLRLEFPFGGDSINGYAH 824
   ||||| : : : : : ||||| : : : : :
Db 999 NNA-----TKLDNMGMA--VOQY IDEQHLGYIYSTNYINADNKL--ANYDAIINAAH 1048
QY 825 DLSLPSIDKGYLTGGOVLAVGTAEVYEFPMKDLRLAVFGD--IGNAYDK--GFTN 875
   ||||| : : : : : ||||| : : : : :
Db 1049 E-----LDK-----VQGNALIAKEAEQKQNIIDAQNALNDQONLANKKXANAFVN 1095

```

RESULT 17  
US-09-815-242-12611

; Sequence 12611, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.

```

; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ. ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12611
; LENGTH: 3158
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-12611

Query Match 2.8%; Score 131.5; DB 10; Length 3158;
Best Local Similarity 19.6%; Pred. No. 3.7;
Matches 187; Conservative 130; Mismatches 337; Indels 301; Gaps 49;

QY 48 AINAK---ACNPPVLTPEIQARL-NAAGLNKPOSALD--VNFDDSPISRIGEQ 101
   ||||| : : : : : ||||| : : : : :
Db 1446 AVNDKNNLHGDKLADOKRATETLNNLSLNT-PQRQALENOINNAATGEVAQKLE 1504
QY 102 SPPGLMSVIEETPLSLELFAOESTEMGINPDYIPE-----YGEQPNSEVVP 154
   : : : : : : : : : : : : : : : : : : : : :
Db 1505 AQALNOMEALRN---SID---QOQTEAG---SKFINEKPOKDAYQAQVAKKLDIN 1554
QY 155 PLEPERKGLIKRLYARLFNDGVNKP---RLKAKFYOSSGSEI-----SAIGSSHOK 205
   ||||| : : : : : ||||| : : : : :
Db 1555 QTSNP-----TLDKAQVEQLQGVNAQAKNLDGDKLADKQHAVTDLNQLNSLNNPQRA 1610
QY 206 TEPRANIKALEITQE---SAMD-----LNGSIPRLQOTALVARAVGYIDISIR 256
   ||||| : : : : : ||||| : : : : :
Db 1611 LESQINNAATREVAQKLAEQALDQAMQALRNSIQOQOTE-SGSKFINEKPOKDAYQ 1669
QY 257 NSIGEVNVIITHDGEPPYIDYRAVEVEGEGADKAFITVADEVPLLIGDFVHGKYE--- 313
   : : : : : : : : : : : : : : : : : : : : :
Db 1670 AAQVQAKDLINQGNP-LDKSOVEOL-----TQAVTTAKNML-----HGDQKLAR 1714
QY 314 -----TKKULIENASAEHGFDFGRMLDRSVYILPD--NTA-----DVSILYDTGTQY-- 359
   ||||| : : : : : ||||| : : : : :
Db 1715 DQQAQVTTVALPN-----LNHAQQAALTDALINAAPTREVAQHVOGTATELDH 1762
QY 360 -----REDEVVFTTIDPKTNOLTTDPDKLPVKRELLEQLTV-----NMGEAVNIQ 405
   : : : : : : : : : : : : : : : : : : : : :
Db 1763 AMETLKAKVQVNTDKAQPNYTEASTD-----KKEAVDQALQAAESTIDPTGNSNANKD 1816
QY 406 AV-BALSNDLIATRYFMNMTETIYFPEREQION--DOVSEGOSSSRTPEAOVDEST-LE 461
   ||||| : : : : : ||||| : : : : :
Db 1817 AVEQALITK--LOEKVVELNGENERVAEKAQAKOTIDOLAHLMNQDITAKONIDQATKIQ 1874
QY 462 PVIEVELTDLIMDISPIEFSASNLIDQKLNLYAKARHLVMDPDRVLAINHDDGVNR 521
   ||||| : : : : : ||||| : : : : :
Db 1875 PIAELVD-----QATOLNSMDQLQQAVERNANVEQ 1905

```

QY 522 SI--LGRISAVNARAILPBDESEN-----EVIDLPERTALA 557  
DB 1906 TVDYTOADSKONAYKQAIA--EAENVLKONSKNOVDQALQNLIMAKQALNGDERVALA 1963  
QY 558 NRKTPADVOSKKVPLVYFVASDPKPRDQIGLWGSPTGRLVTKEHNLINRDGYQAGA 617  
DB 1964 KTNCKHIDQALNAL-----NNNQDQGFKC-- 1987  
QY 618 ELRLSEDKKVKLYATPRLSHPLN-----DQLRATLGYQOEVEFGH-- 657  
DB 1988 --RIDQ-----SHDLNQIQOYDEAKALNRAMQOL-----SOEISGNEGRT 2026  
QY 658 --SUNGNDLSRLE-----HEISRSIIONGNNRRTSLRYRLDKLTKTOAPPEWQOLP 709  
DB 2027 KGSITVYNADTQVQYVDEAVDQAKQALDKSTGQNLTAEOYIKLDAVTAAKKA----- 2080  
QY 710 VDFVNGRP-----SOEALLGAVAHKTVADNLVPMRCYRQRYSLVSSGLVSDANNAI 764  
DB 2081 ---LNGEERLNRKSEALQR-----LDQLTHLNNQRLQALQINNAAETLNKASRAI 2129  
QY 765 ARAGISGVYSGDNAYGSRNRAHQMTGGIOAGYIMSDNFNHVYRLRFRFAGDOSIRGYAH 824  
DB 2130 NRA-----TKLDNAMGA--VOOYIDEQHLGIVISSTNTYNINADNMLK--ANYDNALANAAH 2179  
QY 825 DLSLPSIDSKCYLTGGOVLAVGTAEYNYEFMKDLRLAVFGD--IGNAYDK--GFTN 875  
DB 2180 E-----LDK---VOGNALAKAEQOLKQNTIDAQNALNGDONLANAKKANAFAVN 2226  
RESULT 18  
US-09-915-242-5835  
Sequence 5835, Application US/09815242  
Patient No. US0020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA 011A  
CURRENT APPLICATION NUMBER: US/09/815, 242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 5835  
LENGTH: 2434  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-915-242-5835  
Query Match 2.8%; Score 131; DB 10; Length 2434;  
Best Local Similarity 17.6%; Pred. No. 2.7;  
Matches 181; Conservative 157; Mismatches 346; Indels 342; Gaps 46;

QY 29 LAOONNPANIINHVPADHTALNOAKAGNPVLLPEEOIARLNAAGLNAPQOALDVVN 88  
DB 1176 ISATNPEMNPDTIQKASQVNSAKSA-----LNDE---KLAAKQAKASDIGNLTDLN 1227  
QY 89 FDDOSPISRIEQQSPPLGIDMSVIEETPLSL-----BELFAOESTENGINPNDYIEBY 142  
DB 1228 NAORTANAAEVDQAPNLAAVTAARKKATSLNPMGNLKHALLAEKDNTRSVNYVT-- 1282  
QY 143 QGEOPNSEVVVPPLEPEKPGILIRLRYARLFPNDGVNKPRLKAKFYSSQSESTSAIGSS 202  
DB 1283 -ADQPKQA-----YDTAYTOAEAL-----TANAGSN 1308  
QY 203 HOKTEPYANIKAALEDITOESAMDNGSIPRLQYALYAARAVGY-----D 249  
DB 1309 ANETO-----VQALNLQNLQ--AKNDLNGD--NKVQAQKESAKRALAYSNLNNNAOSTAATSO 1362  
QY 250 ID-----LSTIRNSIGEVDYIHDLGEPPYIDYRAVEVRGEGADKAFVTVADVEPLL 302  
DB 1363 IDNATTVAGVTAAQWTANLNTAMQOL-----ONGINDO--NTVKQOVNFT 1406  
QY 303 IGDVFNHKKYEPRKMLIENASA-----EHGYFPGRWLDSVDVILPD--NTADVSLIYDTGT 357  
DB 1407 DAD--OGKKDAYTAAVTAAGCITLKAHQ--NMTKAQVEALNQVTTAKNALNGDANY 1460  
QY 358 QYRFEDEVVFTTIDPKTNQLTTPDKLPYKRELLLEOL--LTVN-----MGEAY 402  
DB 1461 RQA-----KSDAKANLGTLHLNNAQKQDLTQLEGATTVYNGVYKTKAQDLDGAMQ 1513  
QY 403 NIOAVRALSNDLIATRYFRMNVTEIYFPEREQIONDQVSFEQSSSRTPPAQVDESTLER 462  
DB 1514 RLQSAIANKDQTKASE--NYIDAD--PKTKAFPMALITQAESYLNKDKGANKDKQAVSQ 1568  
QY 463 VTEIYELFDGILMDSPIESASNLIDOKLNL-----VAAKARHLIDMP 506  
DB 1569 AIQSVTSTENALNGCANLORATEALQIDNLTHTLPKTKALKQOVNAAQVSVGTDLK 1628  
QY 507 DDRVLAINHDDGVNRSIL-----GRISAVSAVARAILPBDESENEV 547  
DB 1629 NSATSLNNAMQDLQALADHDITVAGSYNTNAPSPOKQAYTDAVYNAKNIV--NGSPNVI 1686  
QY 548 IDLPERTALANKTRPADV-----YOSKVPILVYFVASKPRD--GQI----- 587  
DB 1687 TNAADVTAATQYVNNNAETGLNGDNLATATAKQQAOKALROMTHLSDKOSITGQIDSAFO 1746  
QY 588 --GLGMSPTGTRL--VTKFEHNLINRDGYOAG----- 616  
DB 1747 VTGQVQSKDNATNLDNANMQLNSTANKQDVKASQPYVDADKQNAVTFVYTNMENIIN 1806  
QY 617 -----AELRLSEDK--KGVKLYATK-----PLSHPLNQLRATLGY 650  
DB 1807 ATSPTLDPASAVTOAANGVSTKTKALNGAONLANKKOETTANINQLSHNNQO----- 1859  
QY 651 QOEVEGHSHTNGFDLSL--RTLEHEISRSI--IONGNNRRTSLRYRLDKLTKTOAPPET 704  
DB 1860 KODLNTQVTNAPNISTVNOVKTKAEQLDQAMERLNGIDK-----DOVKQS----- 1906  
QY 705 WODLPVDFVNGRPQOEALLAGVAHVKTVDNLVPMRCYRQRYSLVSSGLVSDANNAI 764  
DB 1907 -----VNFTDADPEKQTAVANNV--TAENITINQANG-----TANQSQ 1943  
QY 765 ARAGISGVYSF-----GDNAY--GSRNRAHQMTGGIOAGYIMSDNFNHVYRLRFRFAGGO 817  
DB 1944 VEAALSTVTTTQALNGDKKVTDAKNNAQOTLSTL-----DNLNN----- 1983  
QY 818 SIRGYAHSLSIPDSKGYLTG--GOVLAVGTAEYNYEFMKDLRLAVFGDIGNAYDKGFTN 875  
DB 1984 -----AOKGAVTGINQAHYAEVYQALIQTOELNTAM--GMLKNSLN--DK 2026  
QY 876 DTKIGA 881  
DB 2027 DTTTGS 2032



```

; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5229
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5229

Query Match      2.8%; Score 130.5; DB 10; Length 1184;
Best Local Similarity 18.4%; Pred. No. 0.97;
Matches 143; Conservative 131; Mismatches 287; Indels 215; Gaps 35;

QY 74 GLNAKQSOALDVNFDDSPISRIGEOSPPGLDMSVIEETPTLSLELFAOESTEMGI 133
DB 31 GPNGSGKSNITDAIKW-----VLGEQSAK-SLRGSKMEDIIIFSGAHRKQONVAEYQL 82
QY 134 NPNDIPEYQGEQRPSEVVPPTL---EPE-----KPLIKRLIYAR 171
DB 83 RLDNNSKSLYSVE--NEVIYTRRLYRSGESEYYINNDRARLKDIADLFDSGLGEAVSI 140
QY 172 LFNDGVNKVPRLKAIFYOSOGSETSA--IGSSHOKTEPYANIKALEDIT--OESAMD, 227
DB 141 ISQGRVDEI--LNAKPIRROIIESAGVLYKKRKAESLNLKDOTEDNLTVEIDL 198
QY 228 NGSIPLRLQOTALVARAVGYDIDLSIRNSIGEVDT--IHDLGPV---YIDYRAVE 281
DB 199 EGRVEPLKEEAIAIK-----EYKTLSHQKHSDIYVTVHDIOQYTNDRNQDRLMD 250
QY 282 VRGEGADKATTVVADDEVPL-----LIGGVFHGKYEYTKN 317
DB 251 LOGOANKE-----ADKRLSQOIQYKGRHOLDNYESLNYQLVAKATEAFKETTGOIN 305
QY 318 LIENASAEHGYFDGRWILDRSYDVI--LPDNTADVSLIYDTGTQYRFDEVFPTIDPKTNO 375
DB 306 VLEERKKQSEETNARYEEQENLIELLENISNEISEADT-----YKSLSKQKE 355
QY 376 LTTDPDKLPVRELLEQLLYNMGAEVYMLQAVRALSNDLIATRYFNMYNTELVPEPEROI 435
DB 356 LNA-----VIRELEQOLYVSDEAHEKLEIK-----NEYTTLMS-----EQSDV 395
QY 436 QND-----OVSPFOSSSTPEPAQVDESTLEPIYETVELTDGL-----MDISPI 480
DB 396 NNDIRFLKHTIEENDAKKSRL-----DSRLVEYFEQLKDIOQIKTKKEYOQTNKELSAV 451
QY 481 EFSASNLIODKLNIVAAKARH--LY-----DMPDDR--LAINHDD-----GV----- 519
DB 452 DKEIKNIEKDLDTKKAKONEYEKLYQAYRYTEKKMKTIRIDSLATQEEYVTFEFGVKHIL 511
QY 520 ---NRSLIG-----RISDAVSAVARAI--LPRESENEVID-----LPERPAL 556
DB 512 KAKKEIKLGHGVAEIIIDVPSKLTQALETALGASLHVIYIDSEKGRQAIQFLKERNIG 571
QY 557 ANRKTADVYOSKRVPL-----YVFVADSKPR-----DGQIG----- 588
DB 572 RATPLPLNVIOSRVYANDIKSIKAKEANGFISISEAVKVAPEYONIIIGLNTIYDHL 631
QY 589 -----LOWGSDTGRVLYTKFEHNLINNDGYQAGAEIRLSEDKKGVLYATKPLSLPLMDQ 643
DB 632 KHANELARAIKYRTRIYLT-LEGDIYVPGSGMTGGARKSKSILSQKDELT--TMRHQLEDY 689
QY 644 LRAATLGVOEFGVSTNGFDLSTRTLEHISRSIIIONGMNRTSLKRYRLDKLTKQ 699
DB 690 LROTSEFQOFKELIKISDQLSLEYFKSOKHNFLK-----EYVHFEMELDRLTTO 741

RESULT 21
US-09-815-242-12125
; Sequence 12125 Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl E.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 12125
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12125

Query Match      2.8%; Score 130.5; DB 10; Length 1188;
Best Local Similarity 18.4%; Pred. No. 0.98;
Matches 143; Conservative 131; Mismatches 287; Indels 215; Gaps 35;
```

```

QY 74 GLNAKQSOALDVNFDDSPISRIGEOSPPGLDMSVIEETPTLSLELFAOESTEMGI 133
DB 32 GPNGSGKSNITDAIKW-----VLGEQSAK-SLRGSKMEDIIIFSGAHRKQONVAEYQL 83
QY 134 NPNDIPEYQGEQRPSEVVPPTL---EPE-----KPLIKRLIYAR 171
DB 84 RLDNNSKSLYSVE--NEVIYTRRLYRSGESEYYINNDRARLKDIADLFDSGLGEAVSI 141
QY 172 LFNDGVNKVPRLKAIFYOSOGSETSA--IGSSHOKTEPYANIKALEDIT--OESAMD, 227
DB 142 ISQGRVDEI--LNAKPIRROIIESAGVLYKKRKAESLNLKDOTEDNLTVEIDL 199
QY 228 NGSIPLRLQOTALVARAVGYDIDLSIRNSIGEVDT--IHDLGPV---YIDYRAVE 281
DB 200 EGRVEPLKEEAIAIK-----EYKTLSHQKHSDIYVTVHDIOQYTNDRNQDRLMD 251
QY 282 VRGEGADKATTVVADDEVPL-----LIGGVFHGKYEYTKN 317
DB 252 LOGOANKE-----ADKRLSQOIQYKGRHOLDNYESLNYQLVAKATEAFKETTGOIN 306
QY 318 LIENASAEHGYFDGRWILDRSYDVI--LPDNTADVSLIYDTGTQYRFDEVFPTIDPKTNO 375
DB 307 VLEERKKQSEETNARYEEQENLIELLENISNEISEADT-----YKSLSKQKE 356
QY 376 LTTDPDKLPVRELLEQLLYNMGAEVYMLQAVRALSNDLIATRYFNMYNTELVPEPEROI 435
DB 357 LNA-----VIRELEQOLYVSDEAHEKLEIK-----NEYTTLMS-----EQSDV 396
QY 436 QND-----OVSPFOSSSTPEPAQVDESTLEPIYETVELTDGL-----MDISPI 480
DB 397 NNDIRFLKHTIEENDAKKSRL-----DSRLVEYFEQLKDIOQIKTKKEYOQTNKELSAV 452
QY 481 EFSASNLIODKLNIVAAKARH--LY-----DMPDDR--LAINHDD-----GV----- 519
DB 452 DKEIKNIEKDLDTKKAKONEYEKLYQAYRYTEKKMKTIRIDSLATQEEYVTFEFGVKHIL 511
```





QY	55	GNNPVLTLPPQIOARLMAAGLANKPOSSQALDVNPFDDOSPISRIEOSPPLIDMSVIE	114
Dp	582	GAFELIYRNEK-----RMLOEADVLFDMGRSTNAVKGANKRPL-----	620
QY	115	TTPLSTEEELRAQESTENGITPNNDYITPEYQEOPNSEVVPPLLEPEKQGLIKRLYARLFN	174
Dp	621	---KSLSEIILKKGQGRFQMLLCKRVDFSG---RSYIVVGPMLKDECGLPKNMALLEFK	674
QY	175	DGVKVVRLAKAFYOSSOGETSAICSSHQKTEPYAN-IKAALEDITQESAMDNGSIPR	233
Dp	675	-----PHLLSKL---EERGVATTLQAQRMLIEQKSENEWECLOGETTEBYPVLLNRA-PT	724
QY	234	LROTA-----LVARANGY-----DID-----LSTRNSIGF-----V	262
Dp	725	LHKOSIOAFHFKLIDGKAIOILHPLVCSAFENAPFDGQMAVHPLSOEAIAECKVLMLSM	784
QY	263	DVIHHDGEPEYID-----YRAVERGEGADQKAPTVADEVPLL-IGDVFHNHGK	312
Dp	785	NILLPAGKAVAIIPSDQMVGLYLSLEKSGVKGHEKLFSSVNEITTAIDTKELDINAHI	844
QY	313	ET--KKNLLENASAHEGYFEDGRWLDRSYVDILPD-----NTADVSLIYD----	354
Dp	845	RVLQGNIIATSA-----GRMIIRKS---ILPDFIPTDLMRPMKKKDIGVLVDYVHKV	894
QY	355	-----TGQYRFDEVFETITDPKNOULTTDPDKLPIYR--EL	389
Dp	895	GGCITATPELDNLKTGLGFRYATAGISISMED---LTPKQKQWKEAKYEVKKIOQO	950
QY	390	LEQLLTVNMGEAVNLQAVRALNSDLATRYFNMVNTETIYFPEREIONDOVSFEQ----	444
Dp	951	YDQGLLTDQ-ERY-----NKIIDT--WMEVNDKMKEMMTALADKBESFNITVMA	998
QY	445	SSSSRTPEAOV-----DESTLE-PYIEVELTDCIILMDISPIEFSASNLIOD	490
Dp	999	DSGARGSAAOIROLSAMRGMLTKRDPDSIETETIISNFKGLVLEYLFNSTHARGGLADT	1058
QY	491	KLNLYVA--KARHLDPMDRVLIAHHQDVNSLIGRISDVSVAARILPDESE----	544
Dp	1059	AKTANNGYLTKRLIDV-SQNVAVVSDDCGTHGHT-ETTD-IAVGSLEIPELERINGR	1114
QY	545	-----NEVIDLPERFALANRKTPTADYOSKV-----PLYFVASDKPRD-----GOIG	588
Dp	1115	VLEEDVIDPITNELLYADTLIDEBEAKKVYAGIKSITIRTPYCKAPKGCACGYCLN	1174
QY	589	LGMF-----SDPTGRLTATKEEH-----NLINRQGOAAE	618
Dp	1175	LGBEKMYSPEEAVGYAAOISGEPGTOLLRPFHNGGTAASRODERETIVASKEGVFRYN	1234
QY	619	LRLSEDKKGVKYATK-----PLSHRPLNDOLRATLGYQDEFVGHSTNGEDLSTRT	668
Dp	1235	LRTYTNKEGKNITIANRBNASILVVEPKIKAPDGLRITETVEEYVVS-----	1282
QY	669	LEHIEISITIQNGCMNRTYSLRRLDKLKTQAPPELMQDLPVDFPNGRPSOEALL--AG	725
Dp	1283	-----VKNDOEAKFEVLR-RSDIYK--PSE-----LAGVGKIEGKVLPLVYASG	1323
QY	726	VAAVH--TVADNLVYNMGRQRYRQRYSELEVSQGLVSDANNAIAR--AGISGYVSPGNAY	780
Dp	1324	HKVHKGSSTADIT---QEGCMNVNENRIPYASSELLVND-NMPIADQVYAAKEKVIKY-----Y	1375
QY	781	GSNBAH-QMTGCIQAOYIMSDNENHVPYRLRFA-----GGDQSTIRGY-AHDSLSPISDK	833

```

1  GENERAL INFORMATION:
2  APPLICANT: Haselbeck, Robert
3  APPLICANT: Ohlsen, Kari L.
4  APPLICANT: Zyskind, Judith W.
5  APPLICANT: Wall, Daniel
6  APPLICANT: Trawick, John D.
7  APPLICANT: Carr, Grant J.
8  APPLICANT: Yamamoto, Robert T.
9  TITLE OF INVENTION: Identification of Essential Genes In
10 TITLE OF INVENTION: Prokaryotes
11 FILE REFERENCE: ELITRA.011A
12 CURRENT APPLICATION NUMBER: US/09/815,242
13 CURRENT FILING DATE: 2001-03-21
14 PRIOR APPLICATION NUMBER: 60/191,078
15 PRIOR FILING DATE: 2000-03-21
16 PRIOR APPLICATION NUMBER: 60/206,848
17 PRIOR FILING DATE: 2000-05-23
18 PRIOR APPLICATION NUMBER: 60/207,727
19 PRIOR FILING DATE: 2000-05-26
20 PRIOR APPLICATION NUMBER: 60/242,578
21 PRIOR FILING DATE: 2000-10-23
22 PRIOR APPLICATION NUMBER: 60/253,625
23 PRIOR FILING DATE: 2000-11-27
24 PRIOR APPLICATION NUMBER: 60/257,931
25 PRIOR FILING DATE: 2000-12-22
26 PRIOR APPLICATION NUMBER: 60/269,308
27 PRIOR FILING DATE: 2001-02-16
28 NUMBER OF SEQ ID NOS: 14110
29 SOFTWARE: FastSeq for Windows Version 4.0
30 SEQ ID NO. 11410
31 LENGTH: 2890
32 TYPE: PRY
33 ORGANISM: Helicobacter pylori
34 US-09-815-242-11410

```

```

Query Match 2.7%; Score 12.7.5; DB 10; Length 2890;
Best Local Similarity 19.2%; Pred. No. 6.4;
Matches 193; Conservative 147; Mismatches 354; Indels 309; Gaps 52.

OY 55 GNPPLVLPPEQIQARLNAGLNAKPQSQALDVANFDDQSPIRIGESQSPGLDMSVIE 114
| : : : | : : : | : : : | : : : | : : : |
Db 1675 GAPELIVREK-----RMLQEVADVLPDNGRSTNAVKGANKRPL----- 1713

OY 115 TTPPLLEELFAQESTEMGINPDYIPEYQEQDQNSEVVPPTLEPEKPLRLYARLFN 174
| : : : | : : : | : : : | : : : | : : : |
Db 1714 ---KSLSEITIKQGRFRONILGRKRVDFSG---RSYIVVGPMUKMDECGLPKRMALPEK 1767

OY 175 DGVNVPRLKAKAFYSSQSGETSALIGSSHQKTEPRAN-IKAALEDITQESAMDINGSI 233
| : : | : : | : : | : : | : : | : : | : : | : : |
Db 1768 -----PHLLSKL---EERGVATFLKQAKRMIEQKSNEWECELOITTEYPVILNRA-PT 1817

OY 234 LKQTA-----LVAAKAVGY-----DID-----LSIIRNSIG-----V 262
| : : : | : : : | : : : | : : : | : : : |
Db 1818 LHKQSIQAFHPLDKGAKIQLHPVLCVSAFMAFDQDQMAVHPVLSQALAEKVLMLSSM 1877

OY 263 DVIILHLEPVPYID-----YRAVEYRGEGADDKAFTTVADEVPRL-IGDVFHNCKY 312
| : : : | : : : | : : : | : : : | : : : |
Db 1878 NLLPASAQRVAIIPQDQMWLGLYVLSLEKSYGKEGHEKLFSSVNEIITADPTMELOIHAKI 1937

OY 313 ET--KKNLLENASAEHGYPFDGRMLDRSVYIIPD-----NTADVSLIYD----- 354

```

```
Db 1938 RVLDDGNITATSA-----GRMITKS---ILPDEIPTDLNRRPKKKDIDVLVDYVHKV 1987
QY 355 -----TGQYRFDEVEFFETIDPKTNOLTTDPDLPKR--EL 389
Db 1988 GGIGITATFLDNLKTGLFRYATKAGISIMEDI-----ITPDKOKMVEKAEVKIIOQ 2043
QY 390 LEOLLVYNGEAYVLAQVRLASNDLIATRYFMVNTIELVPEREOIONDQVSFEQ-----444
Db 2044 YDQGLLDDO-ERY-----NKIIDT--WTEVNDKMSKEMMTALIQDKGFNSIYMA 2091
QY 445 SSSSRPEAVY-----DESTLE-PIVETVELTGIIIMDISPIEFSASNLIQD 490
Db 2092 DSGARSAAOIROLSAMRGIMTKPDGSIETPIISNFKEGINLVEYFNSTGAKRGILADT 2151
QY 491 KLNLYVA--KARHLYPDPRDLVLAINHDDGVNRSILGRISDASAVARAILPDESE-----544
Db 2152 ALKTANNGYLTKRLIDV--SONVKVYSDDCGTHEGI--ETID-IAVGSLEIPLERIFGR 2207
QY 545 ---NEVIDLPERTALANRKTTPADYQSKV-----PLYVFAVSDKPRD-----GOIG 588
Db 2208 VLEEDVIDPTNEILLVADFLIDEGAKKVEAGIKSITITPTVCKAPKVCACKYGLN 2267
QY 589 LGMG-----SDTGRFLVTKFEH-----NLNRDGYQAGAE 618
Db 2268 LGGKMSYPEAVGVVAAQSIGEPGTQLTTRTFHVGSTASRQDEREIVASKEGFVREFYN 2327
QY 619 LRLESDKKGVKLYATK-----PLSHPLNDLRATLGIQOEFGHSHNGDLSTRT 668
Db 2328 LRTYTKKEGNITIANRKNASILVVEPKIKAPFDELRIEYVEVVS-----2375
QY 669 LEHEISRSIIIONGWMNTYSILRYRLDKLTOAPETWODLPVDFVANKPSEQALL---AG 725
Db 2376 -----VKNDGOEAKFVLR--RSDIVK---PSE-----LAGVGKIEBKVLLPASP 2416
QY 726 VAVHK--TVADNLVNPMPRGVROKRSLEVGSSGLVSDANMAIAR---AGISGVSYFGNAY 780
Db 2417 HKVHKGGSIADII--OEGNNVNPRIPIYASELLVKD--NDPIADQVYAKKGVIKY---Y 2468
QY 781 GSNRAH--QMTGGIOAGYIAGSDNFNHPYRLRFPA-----GGDOSIRGY--AHDLSLPSIDK 833
Db 2469 VLEANLHERTHGIKKGDMSYK-----GLRFVIAADNGREKARHYIARGSEIILIDN 2520
QY 834 GYLTGGQVLAVGTAE-----YNYEPMKDLRLAVFGDIG 866
Db 2521 SEVSTNSVISKPTTNTFKTATWDPYNTPTIADPK-----GKVG 2559

RESULT 25
US-10-107-649-2
; Sequence 2, Application US/10107649
; Publication No. US20030044949A1
; GENERAL INFORMATION:
; APPLICANT: Behr, Regine
; APPLICANT: Sloma, Alan
; TITLE OF INVENTION: Polypeptides Having Gamma-Glutamyl
; FILE REFERENCE: 10157, 200-US
; CURRENT APPLICATION NUMBER: US/10/107,649
; PRIOR APPLICATION NUMBER: 2002-03-27
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 604
; TYPE: PRT
; ORGANISM: Bacillus agaradhaerens
US-10-107-649-2

Query Match 2.7%; Score 127; DB 9; Length 604;
Best Local Similarity 21.1%; Pred. No. 0.61;
Matches 116; Conservative 87; Mismatches 226; Indels 122; Gaps 29;
```

```
QY 190 SSGSETAIGSSHOKTEPYANIKAALEDITQESAMD--NCSIPRLROTALVAARAVGY 247
Db 55 SENNGQTAEQNNGEARVYITGSSA--HPLAAVEGMIDIENG-----TALDAVAVSF 106
QY 248 YDIDLSIR---NSIGEDVYII-HDLGEPVY-IDYR-AVENREGAGAD-----289
Db 107 M---LNVPEYGGIGGGVMLYHDPAGVYSYDREAAPISGN--DDPTGCAVAPGV 161
QY 290 KAFITVAD-----EYPLLIGDVFFHHKIKETKKNL--IENASAEHYFD 330
Db 162 KGMDLIHDNHGELPEWEDYIAPAIERAETGFQVGDIFHOOTGNNAVYILMEEHERQLFPE 221
QY 331 GRMIDRSYVDLIPDNTADVSLIDYGT-----QYRDEVEVFFETIDPKTNOL 376
Db 222 GQALGVNQVLQVEDLADTLRLIQENRSQGFYSGPIGLDLOQOFNTEEDLASIEFQITE- 280
QY 377 TTDPDKLPVKRELLLEQLLTVNMGEAYNLQAVRALSNDLIATRYFMVNTIELVPEREQIQ 436
Db 281 -----PVSAEVEGQIYVGGPSPSSGTVVQAL-----QVADQLDLNDVFPD-EDLP 325
QY 437 NDQVSFEQSSS-----SRTEPAOVDESTLEPVETVELTGIIIMDISPIEFSASNLIQD 490
Db 326 EDFESFGDSQHIYIHLINETKANSTKATYDSRLDT--LGPAPFDDIDHQAALTDQYIQ 383
QY 491 KLN-----LVAARAKHLXMPDDRVLAINHDDGVNRSILGRISDAVAVAR---ALLP 540
Db 384 LDDISFNEITPGDTSELPDPAEADSRHTTHFYIDKEKRMASATISLFEFGSGIYI 443
QY 541 DE---SENEVIDLPERTALANRKTTPADYQSKVPLVFAVSDKPRDQIGLWGSDDTGR 598
Db 444 DGFELNNQMTFSDNPDSINRYEPGKRRTFVAPR-IF-----EEEGPVLGKSPGGR 497
QY 599 L-VTKFE-----HNLINRDG---YQAGAEIRLSEDKKGVKLYATKPLSHPLNDLRATL 648
Db 498 IPAWVFQIMQYHGINDGDPMTLQEAIEAPRFNEEDV--IYLQELPEVYSNLR--NM 555
QY 649 GYQOEVEFGHSF 659
Db 556 GY--SVVGHSS 564

RESULT 26
US-09-820-843A-16
; Sequence 16, Application US/09820843A
; Publication No. US2003039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PRO
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 1616
; TYPE: PRT
; ORGANISM: M. genitalium
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: cytochrome-c accessory protein
; NAME/KEY: misc feature
; OTHER INFORMATION: 9111046097
US-09-820-843A-16

Query Match 2.7%; Score 125.5; DB 9; Length 1616;
Best Local Similarity 19.5%; Pred. No. 3.6;
Matches 111; Conservative 94; Mismatches 180; Indels 183; Gaps 31;
Db 297 NTRSLASSL-----ENMLLENSSEBP-----VIEVKPRRNEVIFNPTKLFHEK 344
```

	Query Match	2.6%	Score 124.5;	DB 9;	Length 3712;
	Best Local Similarity	20.1%;	Pred No. 15;		
Matches	183; Conservative	117;	Mismatches	341;	Indels 269; Gaps 42.
QY	110 SYIETTLTSLSEELFAOESTENGIMPNDIPIPYOGEPKPSVYVPPTLEPEKGLIKRLY	169      :-:::-: -:-  -:			
DG	2132 SVLDEPFSNTLAFFTSOKLNYYDQLADELPVKVLDPNPSVDLS-	-:-:-:-:-kKAN 2181			
ZY	170 ARLENFG--VNKVPRLLAKFYOSSOSGETS-----AIGSHOKTEPVANIK---	213			

```

RESULT 28
US-09-995-542-10
; Sequence 10, Application US/09995542
; Patent No. US20020127647A1
; GENERAL INFORMATION:
; APPLICANT: Shutter, John
; APPLICANT: Ullas, Larni
; TITLE OF INVENTION: ATP-Binding Cassette Transporter-like Molecules and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 00-658-A
; CURRENT APPLICATION NUMBER: US/09/995,542
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 60/253,520

```

;; PRIOR FILING DATE: 2000-11-28  
;; NUMBER OF SEQ ID NOS: 24  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 10  
;; LENGTH: 2310  
;; TYPE: PRF  
;; ORGANISM: Mus musculus  
US-09-995-342-10

Query Match 2.6%; Score 124; DB 10; Length 2310;  
Best Local Similarity 19.3%; Pred No. 8;  
Matches 166; Conservative 119; Mismatches 307; Indels 270; Gaps 43;

OY 112 IEETPLSLEELFAOESTENGINPNNDIPEYGEOPNSEVVPPTLEPEKPGI-----IK 166  
DB 894 LEKTEPLT-EEMEDPEHPE-GNMDSPF-----ERELPGLVPGVCVK 932  
OY 167 RLYARLFNGVKNVPRLLAKKFVOS-----SOSGETSAIGSHOKTEPYANIKALED 218  
DB 933 NLVKEFEPSPRAVDRLNITFEYENQITAFLGHNAGKTTTLST----- 975  
OY 219 ITQESAMDNGSIPRLROTALVARAVGYDIDLSIIRNSIG---EVDYIHLGEPYI 275  
DB 976 -----LTGLLPRTSGVYLIGKDI---ETMLDVYROSIGMCPQHNLIFHHLTVAEHI 1024  
OY 276 DYRAVEVNGEGADKAFITVADEPRLIGDV-FHNGKYETKKNL-----IENASAEHG 327  
DB 1025 LEFYA-OLKGRSMEAOI-----EMEAMLEDYGLHKKRNEAOQLSGMOKKLSVALAFVG 1078  
OY 328 YVDGRMLDSVYIILPDNTADVS-----LIYDTGTYRDEVEVFFITIDPKTNOLTTDPK 382  
DB 1079 -----DSKV-VLDEPTSGVDVPSRRSIWDLILKRYSGRTIMS-----THHME 1122  
OY 383 LPVKREILEQLITVMNGEAYNLQAVRALNSDLATRYFMMVNTEIYEPEREQIONOVSF 442  
DB 1123 ADL-----LGRIRALISGRKXCSTPLFLNCRGTGYLTVR-----KKNKIQSOGGC 1173  
OY 443 EOSSSSRFE-----PAOVDESTLEPVI-----ETVELTDGIMDISPIEFSSANLI--- 488  
DB 1174 EGVCSCTSGKSTRCTPTRYKDEITEEQVLDQVQELMDVLYHHNPEAKLVCEIQEILFLL 1233  
OY 489 -----QOKNLIVAAKARHLVMDPDRV-LAINHDDGVNRSILGRISDA 530  
DB 1234 PNKNFKORAYASLFRLEETLADLGSSFGISDTPLEEIFLKVTEPAGAGSMFVG---GA 1290  
OY 531 VSAVARAIL-----PDESENEVIDLPERITALANKRTPADVQSKVPLVYFVASDKPRD 584  
DB 1291 QOKRQOAGLRHCSAPTEKLRQYAOAPHTC-----SPGOVDPPKGP-----SPEPD 1338  
OY 585 GOIGLGKMSDTGTR-----LVTKFEHNLINRDGYOAGAEILRLSEDKGKGLYATKP 635  
DB 1339 P-----GVPEFTGARLILQHVQALVYKRFHHTIRSKDFVQAOIYLPATFYFLALMLSLIYP 1394  
OY 636 LSHPLNDOLARILGYOQEVFGHSTNGFDLSTRTLEH-ELISRTION--GGMNFTYSLRYR 692  
DB 1395 ---PGEERPALTL--HPWMYGHQYTFEFSMDEPNEHELVEVLADVILRRPFGNRC----- 1443  
OY 693 LDKLTQAPPE-----TWODLPYDFVNGKPSQEALL---AGVAHVAKIV----- 732  
DB 1444 ---LKEEMLPEIYPCINATSMKTPSY-----SPNITHLFKOKWTAHNPSPCKSCSTREKL 1495  
OY 733 -----ADNLVPMRNGYRQYSLLEVSSGLVSD---ANMAIARAGISGVYSGDNA 779  
DB 1496 TMLPECPBGAGGLPPQRFQSTEVLDLTNNINISDYLVKTYPALRRSLKSKFWNDEOR 1555  
OY 780 YGSNHAHOMTGIOAGYLSNDNFNHPYRLRFPAGGDSIRGTAHDSLSPISDK---GYL 836  
DB 1536 Y-----GGISIG-----GKLP-----AIPISGEALVGL 1579  
OY 837 TG-GOVLAVGTAEVNYEPMKOL 857  
DB 1580 SGLGOMMNVSGGPVTRASKEM 1601

RESULT 29  
US-09-801-368-364  
; Sequence 364; Application US/09801368  
; Patent No. US20020128250A1  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Call, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: Royer, John  
; APPLICANT: Salama, Sofie  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeff  
; APPLICANT: Summers, Eric  
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
; FILE REFERENCE: 109272.147  
; CURRENT APPLICATION NUMBER: US/09/801,368  
; PRIOR FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 09/487,558  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: US 60/160,587  
; PRIOR FILING DATE: 1999-10-20  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 364  
; LENGTH: 1250  
; TYPE: PRF  
; ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-364

Query Match 2.6%; Score 123.5; DB 10; Length 1250;  
Best Local Similarity 20.0%; Pred. No. 3.4;  
Matches 209; Conservative 138; Mismatches 363; Indels 337; Gaps 58;

OY 33 NNP-----ANINNVPAHDRAINOAKGNPVLTPPQIQARLNA--AGLNAPK 79  
DB 140 NNGGNSHRTSSQSSITYGHSRRHSGLNKAkra-----AAEQAKRISGEGAVYVK- 192  
OY 80 QSOALDVNFDQSPISRIEQS-----PPLGLDMVIEETPLSL----- 120  
DB 193 -----IDSVQADSS--NSTEQSDFKPPPNAAHQHRRATSNLSPSEKFPNNSHCND 246  
OY 121 EELFAQEST--EMGINPNNDIY-----EYQGEOPNSEVVVP-----TL 157  
DB 247 DEFIAITSSTHRBSKTRNNEYSPCINSWMRQSOQPQO--LSPFRHSGNSRDYNSFNTL 304  
OY 158 EPEKPGILIRLYA-RUFNDGVKNVPRLLAKKFYQSSQGETSAIGSHOKTEPYANIKAL 216  
DB 305 EP--PAIFQGHKHRANSNSVHSF-----SSQGNNGN--GGRKSLPAPYLP-QANI 350  
OY 217 EDTQESAMDNGSIPRL-----RQALVARAVGYDIDLSIT-----RNSIGEVYDIHD 268  
DB 351 PELIQGR--LVAGILRVKKNRSDAMVSTD--GALDADYITGSGDRNALBGLV--- 403  
OY 269 LGEPRYIDYRAVEVNGEGADKAFITVADEPRLIGDVFNHNGKYETKKNLLENASAEHG 328  
DB 404 -----AVEL-----LVVDVWESKKEKEKRRKRAASQHD 435  
OY 329 F-----DGRMLDSVYIILPDN-----TADVSLIYTGTYRDEVEVFFITIDPKTNOLTT 378  
DB 436 IPLNSSDDYHNDASTAAATSNMFLSSPSSSDSLSKDOLSVR-----KRSTINN 485  
OY 379 DPDKL--PVK-----RELLEQLITVMNGEAYNL--QAVRALNSDLATRYFMMVNTEIY 429  
DB 486 DSDSLSSPTKSGVRRKSSLKQPTQKKNDDVEYEGOSLLIVEEELINDKTKPLAGHAV- 544  
OY 430 PEREQIONDOVS-----FEQSSSSRTEPAOVDESTLEPYETVELND-----GILM 475

```
Db 545 AYLDRIIGOLFSGTGLLRPSQOANSNNKRPQ-----SPKIAMFKPTDKKVLPIAIP 598
QY 476 DISPIER-SASNLIDQKLNVAARHLYPDPRVLAINHDDGVNSILGRISDAVASAV 534
Db 599 ELAKRDEVENADKSEKLFVASIKRWPITSL-----HPFGILVSELGDIDHPDTEI 649
QY 535 ARALIPDES--ENVVID-----LP-BRTALANKTPADVYQSKKVLVYVA 578
Db 650 -DSLIRNNFLSNRYLQOKNPKPKPSPFQPLPLAESLEYRRNETDINEVN-----IPAI 703
QY 579 SDRPRDQOIGLWGS-----DTGTRLVYKFEHNLINRDGYQAGAEILRSED 624
Db 704 SE-----LGWVSEFALHYRNNNGTLELGCYVVDVTSI-----IEGSSVDRRAR 748
QY 625 KKGKYLATKPLSHPL-----NDQLRALYQOEVEFGHSTNGFDLSTRTLL-----EHE 672
Db 749 KRSAVMPKLVNLLPQSFNDELSTLAPGKESATL---SVYTTLDSSYLRIRKSTWVGEST 805
QY 673 ISRSIIIONGCMNRTSYRLYRDLKLTOAPP---ETWODLPVDFVNGK--PQOEALLAGVA 727
Db 806 ISPSNI-----LSLEQDEKLSTGSPITSYSTVQEIANSFYARRINDEPATILLPTLS 857
QY 728 VHKTVAD-----NLVPMRGY-----RORYSLEVSSGLVSDANMAIAR 766
Db 858 LLESIDDEKVKVDNITLDRTLGFVYINERKRVNSTVAEKITYTKLGLDALLRQWQPIAT 917
QY 767 AGIS---GVYSFGNAXGSNRHAHOMTGIO-----AGYI 797
Db 918 KMASFRKKIYFEGYN-FDTNPADELIGVLIKDDVAVGIEILLFKTMPRARRYFIACKV 976
QY 798 WSDNFNVHVPRLRFAGAGDSIRGVA---HDSL-SPISDKGYLLGGOVLAVGAEVYVE 852
Db 977 DPQGYHVALNLPITYHTAPMRKRYADHVYRQAKAVIHDPYTEDMALAI-TSEY-CN 1034
QY 853 FMKD-----LRLAVFGDIGN 867
Db 1035 FKXDCAYAOEQAIHLLCKTINDMGN 1061

RESULT 30
US-09-738-626-6999
; Sequence 6999, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738.626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6999
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6999

Query Match 2.6%; Score 121.5; DB 9; Length 852;
```

```
Best Local Similarity 21.0%; Pred. No. 2.6;
Matches 193; Conservative 113; Mismatches 318; Indels 295; Gaps 51;

QY 1 MSKPYLFANRSEFMPALAAVPLMTSOALAOQNAN--IINHVAHDPAINQAKGNPV 59
Db 45 VAAPPLMATGVDPKELILAEAKKLVASYPKASGAMNANPINFNDALNAFTAQELAGE--- 101
QY 60 LITPEQIARLNAAGLNAKPOSQALDV-----NFD----- 90
Db 102 -LGDEYSTEVLLAGI-ARQSDAADLLTNKGATYDAIKKAPPSVRSQRYTTODPEQF 159
QY 91 -----DQSPISRIQSGPPLGDMVIEETPTLSLEELFAQSTENGKINPNDIPEY 142
Db 160 QALEKYSTDLTKLAEKQIDVPYIGRQDEIRRYOYLSR-----TKNN--PVL 205
QY 143 OGEOPNSEVVPPTLEPPEKPGILKRLYARLFENDGVNKPV-LKAFYOSQSGETSATGS 201
Db 206 IGEPGVKTAI-----VEGLARRIVA-----GDVPESLKGKTLTSLDYG--SMVAG 249
QY 202 SHQKTEPYANIKAALEDI-----TOESAMDUNGSIPLR---RQ 236
Db 250 AKYRGEFEERLKAVIDEIKANGAEVYTFIDELHTTVGAGASGSEAMDAQNMKPLIARGE 309
QY 237 TALVAAARAVGYDDIDLSIIRNSIGEVDYIHDGEPVYIDYRAVEVRGEGADKAFITVA 296
Db 310 LRLVQATTLINEY-----RKTI-EKDAALERFRFOVY-----GE-----PYVE 346
QY 237 DEVPLIG---DVFHGKYTEKKNLIEMSAEHGYEDGAWL-DRSYVDILPDNTAD-V 349
Db 347 DAIGILRLKERIEYV-HHGVRIDQSALVAABELSRITYTSRFLPKAIDLV--DEAASRL 403
QY 350 SLIYDTGYQYRDEY--VEFTIDPKTNOLTTDP-----KLPVKELLEQLLT 395
Db 404 RMEIDSSPO-EIDELERIVRLIEEMALSKESDAASKERLEKRLSELADEKEREKSELKA 462
QY 396 VNMGEAVYMLQAVRALSDLIATRYFNMTETIV-----FPEREQIONDO 439
Db 463 RMONKETAIDVREKKELEALR-----SESDLAENDQNGRAVELRYGRIPLEK-----Q 514
QY 440 VSFQSSSSRTPEACV-DESTLEPIYETVELTDGILMDISPIESASNLIDQKLNVAAK 498
Db 515 IEDASKYEVENNMLTDEEVPPTIADVVASMTGI-----PAGKMMQ-----GE 558
QY 499 ARHLYDMPDDVLAINHDDGVNSILGRISDAVASVARAILPDESENVIOLPERTALAN 558
Db 559 TEKLINM--ERYLG-----NR-VVGOL-EAVTAVSAV--RRSRAGVAD-----PN 598
QY 559 RKTADVYQSKKVPPLYVAVASDKPRDQOIGLWGSDTGTRLVYKFEHNLINRD----- 611
Db 599 RPTGS-----FLFL-----GPTGVG-----KTELAKAAVEFLPDDDRAMIRID 636
QY 612 -----GYQAGAEILRSEDK--GVKLYATKPLSHPLINDQLRATL 648
Db 637 MSEYGEKHSVARLVGAPPGYVGDGGOLTEAVRRPTVYLDFDEVEKAHP--DVFDITLL 694
QY 649 GYOQEVFGHSTNGFDLSRTLEHE-----ISRSIIIONGCMNRTSYRLYRDLKLTOAPPET 704
Db 695 QVDE--GRLLDG---QGRIVDFRNTIILTSNLGAGG-----TREQMDAVKMAFKFBEF 744
QY 705 WQDLPVDFVNGKPSQEALLAGVAHV-KTVADNLVPMRGYRORYSLEVSSGLVSDANMA 763
Db 745 VNRLDDVVIYFDRLSPQEQITSLVIDIQKQLTDRLAG-----RLNLRVSDS-----AKAW 793
QY 764 IARAGISGVISFGDNAGS 782
Db 794 LAERGY-----DPAYGA 805

RESULT 31
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
; US-09-740-274-6

Query Match      2.6%; Score 121; DB 10; Length 1430;
Best Local Similarity 18.8%; Pred. No. 6.3;
Matches 216; Conservative 160; Mismatches 396; Indels 378; Gaps 59;

QY 25 TSOALQONNPANIINHVAHDTAINOAKA-----GNPVLITP-----EIOIA 68
DB 124 TSEERTKQSEEA-----QTASSOALQAKAKELTKQRTAAQENKKNVDLAIPNVQIOG 179
QY 69 RLNAAGLAKPOSO-ALDVN-----FD-DQSPISRIGEOSPLGIDMSVIEETP----- 117
DB 180 KYYTIGSDQPKRKNAFLTVNNKVLXFYDKNTGALTDSQYQFGKL-TKLNNDYTPHNQIV 238
QY 118 -----LSLEELFAQESNEMINDYIPEYQEQPNSEVVPPTLE-----PEKGLIKRLA 170
DB 239 NFEINTSLEITIDYIVTADSYTRPKDILKNGKTWTASSESLRPLKMSWWDKQTOT--AFL 296
QY 171 RLFDN-----GVNKVP-----RLKAFYQSSQSCETSAIGSSHOKTEPY 209
DB 297 NYMNOQGLGTGENYTAADSSQESLNLAQTVOYKIEIKISQITQOTQMLRIINSFYVTTQP- 355
QY 210 ANIKALEDIQESAMDINGSTIPRLROTALVAAA--VGYIDIDLSIR----- 256
DB 356 -NMNSQTESDSAGKED-----HLOGGALLYSNSDKTAYANSVDYRLNRTPTSQTGPK 408
QY 257 -----NSIGEVDIIT--DLGEVY-----YIIOYRAV-----EVRGGADDKATF 293
DB 409 YFEDNSSGCTPFLANDINDNSNPVQAEOLNWLHYLMANGSIVANDPEANFGVRAVD 468
QY 294 TVADEVPPLIGDVF--HHGKYETKKNLIENASAEHGYFDG--RWLDRSVDVLP--DNFAD 348
DB 469 NVNADLLQIASDYKKAHGYGVKSEKNALNHLISLEAMSDNDQYKDKDTGAOLPFDNKR 528
QY 349 VSLIY----- 358
DB 529 LSLIYALTRPLEKDAKSNKREISGLEPIVTNLSNRSAGKSEMANIIFIRADSEVQ 588
QY 359 YRFDEVVFTIDPKTNOUETDQDKLPV-----KRELLBOLLVNMGEAVNDAVRL 410
DB 589 TVIAITIAQINPKTDGLFTLDELKQAFKIYNEDMROAKKKYTSNIPYAVALM-----L 644
QY 411 SNDLIATP-----YFNMYNT-----EIVFPEREOIQ 436
DB 645 SNKSITPFLLYGDMYSDGQYMATKSPYYDAIDTLKARIKAAAGQDMKITIYVGDSH 704
QY 437 ND-----QVSFEQSSSSRTEPAQVDESTLEPIETVELDGLIMDISP 479
```

```

DB 705 MDMDYTGVLTSVRYGTGANEATDQGEATKTQGMAYITSN-NPSLK-LANQDKVIYVMGA 762
QY 480 IERSASN-----LIDOKLNLVA-----AKARHLXMDPDR-VLAINHDDGVNRSITGRISDAY 531
DB 763 AHNQOEYRPLLTTKDGLSYTSDAAKSLYRKTNDGELVFPASD-----IGYLNPOV 817
QY 532 SAVARAILP-DESENEVIDPERTALANR-KTPADYQSKKV-----PLYVFA 578
DB 818 SGYLAWVPYGASDNDV-----RVASNKANATGOYIESSALDSQILYGFSPFQDFVT 873
QY 579 SDK-----PRDQIGLNG-----SDTGFRLVTFEHNLRNDRGYQAGAE 619
DB 874 KDSDYTNKKIAQNVQLFKSGWGVTSFEMAPQYVSESDGSFDSITQNGYAFEDRY-----DL 929
QY 620 RUSED-----KKGVKIVATKRLSPHNDQLRATIGYQOEYVGHSTNGF 662
DB 930 AMSKNNKYGSOQDMINAVKALHKSIGIOWADW-----VPOIYNLPG--KEV-TYTRVN 981
QY 663 DLSTRTLEHISRSII-----ONGWNNRYSRLRYR--LDKLTQAPPETWODLPVDFVN 714
DB 982 DYGEYRKDSIKKTLVLAANTKSG--KOYQAKYGGAFSELAAKYPSIFNR--TQISN 1035
QY 715 GK-----PSQEL-----LAGVAVHKTYADNLVN--PMRGYR--QRYSLGVGS 753
DB 1036 GKIDPSEKITAWKAKYFNGTNIILGRGVYVLKDNASDKYFELKGNQTYLPRKQMTNKEAS 1095
QY 754 SGLVSDAN--MAIARAGISGVYSFGDNAYGSNRAHQTGCIAGCIWSDNFHVPRLRF 811
DB 1096 TGEVNDNGMTEYSTSGYQAKNSFQDAGKN-----WYFDMNGHMVY----- 1138
QY 812 FAGDQSIGYAHDSLPSIDKGYLTGGOVLAVGTAEVNEFEKDLRLAVFGDIGNAYDK 871
DB 1139 -----GLOQLNGEVOYFLS-----NGVOLRESEFLENADSGKNY-----FGHLGRYSN 1181
QY 872 G---FTNDTK 878
DB 1182 GYYSFNDNSK 1191

RESULT 32
US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LOBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
; US-09-995-749A-2

Query Match      2.6%; Score 121; DB 9; Length 1781;
Best Local Similarity 18.0%; Pred. No. 8.8;
Matches 173; Conservative 119; Mismatches 326; Indels 344; Gaps 44;

QY 133 INP-----NDYIPEYQEQPNSEVVPPTLEPEKPGILKRLYARLFDGVNKPRLAKFYQ 189
DB 668 VNPALSNSSMSGFG-----ITLTPVYKNEVNUQLVHR-----FSN 703
QY 190 SSQSGETSAIGSSHOKTEPYANIKALEDIQESAMDINGSTIPRLROTALVAAAAGVYTD 249
```

```

Db 704 DVKGBEN-----YDFWSELMPPV-KDSFQKNGP---LKQGLGTINGQOQY 747
QY 250 IDLSIIRNSICE--VDYIHLGEPVYIDYRAVEBREGADAKAFYVADVPILLIGVF 307
Db 748 -----LPTTGQPKRKNLLQSGNMWTFD-----SDTGVGTALQLQAKGIVS 791
QY 308 HHGKYEKKLLENASADHGYFDGRWLDKRSYDVLPPDNTADVSLYDTGTQYRDEVV-- 365
Db 792 SNEQY-----RNGMAYSYD-----DKSIEVNGYLLAD-----TWYRPKQILKD 831
QY 366 --FTIDPKT-----NLT-----TDPDKLPVKR 387
Db 832 GTTWTDSKETDMRPIILMMWPNLTLQAYLTYKMKOHGMLPALSPLFENADAPALNHS 891
QY 388 ELLEBOLLTVNGEAYNLQAVRALSNDLIATRYFMMVNTETVEFPEREQIONDOVSFEQSS 447
Db 892 EIVQONIEKRISETGNTDMLRTLMHDFVTNNPMKNKSENVNFGIOQOGFLKYEUSD 951
QY 448 S-----RTEPAQVDESTL-----EPVETVELT----- 470
Db 952 TPVANSOYRLLRMPRIKQOTYRQGFLLANDIDNSNPVQAOQLMWLYLLNFGTITA 1011
QY 471 -----DGLIMDI-SPIEFASNLIDOKLNLV-----AAKARHLYDMPDRLVLAINH- 515
Db 1012 NNDQANFDSYRVADPMDIDADMIAADYFNAATGMSDAVSNKHINLEP-----WNHA 1066
QY 516 -----DDGVNSILGRISDA-----VSAAVARAILPDESENEVIDPERTA 555
Db 1067 DPEYFNKIGNPQMLTDITIKNSLNHGLSDATNRWGLDAIVHQSADRENNST----- 1118
QY 556 LANKRTADYVQSKVPCLYVYFV--ASDKPRQSGIGLWGSOTGRVLYKFEHLNLRQCY 614
Db 1119 -----ENVYIPNTSFVRAHDNNSODQI-----ONALDYGKDIHTTFEEOQ 1161
QY 615 AGAELRLSEDKKGVKLY--ATKPLSHPLNDQLRATLG--YQOEYF-----GHSTNGFD 663
Db 1162 KGIDAYIODQNSYKRYNLVNPASVAILLTNKTOTIPRYVGYDGTGGQYMEHQTRYD 1221
QY 664 LSTRLEHEI-----SRSTIONGNNRTY--SLRY-----RLDKIKTQAPPETW 705
Db 1222 TLTNMLKSRKYVYAGGQSMQTMVSGNNMILTSYRGKAMTATDTGTDERTGIGVVV 1281
QY 706 QDLPEVDVNGKPSOELLAGVAVHK-----TVADNLVN-----PMKRGORRSL 749
Db 1282 SNRP-NIKLGVNDKVVYAHMG-AAHKNOQYRAAVLTITDGYINTYISDQAGAPAMDEMGL 1339
QY 750 EVGSSGLV-----SDANNAI-----ARAGISGVYSFGDNAY 780
Db 1340 YLSSHNLVYNGKEADPAVOGYANPDVSGYLAVVYPVGCASDNODAPRSTKNSGNSAY 1399
QY 781 GSNRAHQMTGIGIAGYIWSQNFNNHVPYR-----LRFAGDQISRGIAHSLS 828
Db 1400 RTMAAFDSNVIFEA-----FSNFYTTPTKESERANVRIAQNADEFAS-----LGFSTSEMA 1450
QY 829 P-----ISDKCYLTGGQYLAAGTAENYEFMKDLRLAFVCGDIGNAYDKFTNDRTIGAGV 884
Db 1451 PQYNSKSDRFLD-----STIDNGYAF-----DXYDLGMBEPKNYGTIDED 1491
QY 885 VR 886
Db 1492 LR 1493

```

RESULT 33  
US-10-060-036-71

; Sequence 71, Application US/10060036  
; Publication No. US20030073144A1  
; GENERAL INFORMATION:  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Persing, David H.

```

; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1192
; TYPE: PR
; ORGANISM: Homo sapiens
US-10-060-036-71

```

Query Match 2.5%; Score 120.5; DB 9; Length 1192;  
Best Local Similarity 18.2%; Pred. No. 5.2;  
Matches 120; Conservative 102; Mismatches 245; Indels 193; Gaps 30;

```

QY 2 SKPVLFANRSPVALAAYLPLMTSQALAOQNNPANTINHVPAHDTAINQAKGNPPLL 61
Db 449 STPEGIKDRSGAYITCAFPNPAAT-ESIA-----TNIFPLGDPTSEKTKTEKIEE 499
QY 62 TPEOIGARLNAGINAKP-----OSQALDVYVPDQSPISIRIGESPLGIDMSYIEETT 116
Db 500 KKAQIVTEKNTSTKTSNPFVLAADSETDYVTTDLNLYTEEVYANMPEGLTPDLVQAC 559
QY 117 PLSEEL-----FAQES-----TEMGINNDYIPEYQGEOP 147
Db 560 ESEINEVYGTKIAVETKMDLVQTSFVMOESLYPAQOLCPSEESFATSPPLDIMEAP 619
QY 148 -NSEY-----VVPPTLEPEKPGIKRIYARLFDNGVKKYRLAKRFQSSQSGTSAIG 200
Db 620 LNSAVPSGASVIOPSSPLEASSVN--YESIKHEPENPP-----YEEAMSVLKKVS 671
QY 201 SSHQTEPYANIKALIEDITQ---ESAMDLNGSIPRLQTAALVAVAGYDI-DLSIIR 256
Db 672 GIKKEIKPENINMALQETEARYSISICDL-----IKETLSAPADPFDSDYSMAKVE 725
QY 257 NSICEVDYIIDL---GEPIYIDYRAVEBREGADAKFTTV---ADEVPILLIGDVPHHG 310
Db 726 QPVPDHSSELVEDSSPDSEPVLD-----FSDSIPVPOKQDFTVLMVXESLRET 774
QY 311 KYET-----KNLEMSAHEG--YFDS--RWLDRSVVILLPDMTADVSLYDTGTQYRF 361
Db 775 SFESMIETENKISALPPEBGKPLLESFKLSLNDTKTLPLDEVSTLS----- 823
QY 362 DEVEFTIDPKTNQITDPPDKLPVKRELLEQLTVNMGEAYNLQAVRALSNDLIATRYFN 421
Db 824 -----KKEKIPIQ---MEELSTA-----YISND----- 843
QY 422 MVNTEIYFPEREQIONDOVSFEQSSSKTEPQVDESTLEPIYETVELTGLIMDISPIE 481
Db 844 ---DLFISKEAQIRETE--FSDSS-----PIETIDFEPPLISKST 879
QY 482 FSASNLIQDKLNLVAAARHLYDMPDDR-----VLAINHDGVNSILGRISDAY----- 531
Db 880 DSFKLAREITDLVSKSELIANAPDGAISLPCTELPHDLST-KNIQPKVEKISFSDDF 938
QY 532 ---SAVARAIL--PD-----ESENEVIDLP--RTALANRKPADYVQSKKVPCLYVFA 578
Db 939 SKNGSATSKVYLLPDPVALATQAEISIVPKVLYVKEAKKLPSDREKEDRSPSAIFSA 998

```

RESULT 34  
US-09-758-140-6

; Sequence 6, Application US/09758140  
; Patent No. US20020012965A1  
; GENERAL INFORMATION:  
; APPLICANT: Strittmatter, Stephen M.  
; TITLE OR INVENTION: No. US20020012965A10 Receptor-Mediated Blockade of Axonal Gro  
; FILE REFERENCE: 44574-5073-US  
; CURRENT APPLICATION NUMBER: US/09/758,140







```

US-09-815-242.12151
Query Match      2.5%; Score 119.5; DB 10; Length 1168;
Best Local Similarity 18.8%; Pred. No. 5.9;
Matches 169; Conservative 116; Mismatches 279; Indels 333; Gaps 42;

QY 108 DMSVIEETPLSLSELEFAOESTENGINPNDI-----PEYOG-----OPN 148
DB 346 DPELLLEGYVTVYFSLSFA-----TPMRPLNHIKRSCKPVQOFGYQYDILMRSEFORVYVN 402
QY 149 SEVVVPETLEPEKPGLIKRLYLARLENDGVNKPRLAKKPYOSSOGSETSAIGSSHOKTEP 208
DB 403 YHIVLVATEETK-----VERQAMLESEMHPSI-----TKLHRSMSSGQAVITIGS--LSEG 452
QY 209 YANIKALEDIQOESANDLNGSIFRLQOTALVARAV-GYYDIDLSITRNSIGEDVYI 267
DB 453 FELPDMLVITEELFEKSKOKQRKPKAISNEKIKSYODLVNG-----DYIVH 503
QY 268 DLGEPYVIDRAVEVRGEGADKAFITVADEVPILLIGDFHNGYETKKNLINENASAEHG 327
DB 504 -----YHNGV----- 508
QY 328 YEDGRWLDRSVYDILPNDTADVSLIYDTGQYREDEYVFTID--PKTNQTLTTPDPLRV 385
DB 509 ---GRYLGVEITLEVQTHRDYIKL-----QYKTDQLFVVDMDQYQKVVASBDKPK 559
QY 386 KREL--LEQLLYNMGSAYNLQAVRALSNDLIATRYFNMTVELYVFPREDIQNDQVSFE 443
DB 560 LNLIGSEEMKT-----KAKQOQSVEDIAEELID-----LYKEREM----- 595
QY 444 QSSSRPEAQVDESTLEPIEIVELDGLIMDISPIEFSASNLQOKLNLVAAKARHLY 503
DB 596 -----AEGYQYEDTAEQ--TFEL--DPEYELTPQOASIDEIKDDMO---KSR--- 638
QY 504 DMEDDRYLAINHDGVNRSILGRISDAVSARAILPDESENEY----- 548
DB 639 --PMDRLLC--GGVG-----YKTEVAVRAKFAVMGKQVAFVPTTILAQOYETLIE 689
QY 549 ---DLBERTALANR-KTPADVOSKKVPLVYFVASDKPRDQIGLGWSDTGTRLVYKFE 604
DB 690 RMODFPEIQLMSRFRTPKEIKQYKE-----GLKTFVYDIVGT 728
QY 605 HNLINRD-GYQOAAELRLSDKK-GVK-LYATKPLSH-----PLNDQLATLIGYQ 651
DB 729 HKLLSDIOYKDLGLLVLDEOQRGVANHKRIKTLKHNDVLLTATPIPTLMSM-- 785
QY 652 QEVFGHSTNGFDLSTRLTLEHISRSIIQONGWNR---TYSLRRLDKLTKQAPETWOD 707
DB 786 -----LGVDNL-----SVIEPPENKFPQYTVLEBNMSFIKALERELSD 827
QY 708 LPVDYNGKPSQOELLALGAVAHKTVADNLVNPARG-YRQRYSLIEVGSSGLVSDANMAIAR 766
DB 828 GGVFY-----LYNKVQSIYEKREQLOM---LMPDANIAVAH 860
QY 767 AGIS-----GYVSEDDNAY-----GSNRAHQWTCG 791
DB 861 GGMSEDRLEETMSLSEINNEEDILLVTTIITETGVDPNANTLIIEDADRFLGSQLQYQGR 920
QY 792 I-----QAGYIWSDFNHHVRYRLFFAGGQDSIRGVYAHDSLSPISDKGYLGGVGLVAGTA 847
DB 921 VGRSSRLIGAT-----FLHPANKVLVETLEDRLQAIKEFEELSGKRIA----- 964
QY 848 EYNVEFMKDLRAVEGDIIGNAYDKFTNTDKIGAGGVNRPVGGVRYVDATGVKE 904
DB 965 -----MRDLNIRGAGNLLGKQOHHFI-DT-----VGFDLYQLMEBAVNEKKGIKE 1009

RESULT 39
US-09-789-386-2
; Sequence 2, Application US/09789386
; Patent No. US20020010324A1
; GENERAL INFORMATION:
; ABPLICANT: MICHAEL OVICH

```

```
APPLICANT: PRINJHA, RABINDER KUMAR
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GP-30165-C1
CURRENT APPLICATION NUMBER: US/09/789,386
CURRENT FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: U.K. 9916898.1
PRIOR FILING DATE: 1999-07-19
PRIOR APPLICATION NUMBER: U.K. 9816024.5
PRIOR FILING DATE: 1998-07-22
PRIOR APPLICATION NUMBER: US 09/359,208
PRIOR FILING DATE: 1999-07-22
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1192
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-789-386-2
```

```
Query Match 2.5%; Score 119.5; DB 10; Length 1192;
Best Local Similarity 18.0%; Pred. No. 6.1;
Matches 116; Conservative 99; Mismatches 241; Indels 187; Gaps 29;
```

```
QY 19 AYLPMTSQLAQQNNPANIINHVAHDIAINQAKAGNPVLLTPQIOARLNAAGLNAAK 78
DB 460 AYTCAPFNPAATESIATNIF---PLGDPITSEKNTDEKKIEKKAQIYTEKNTSTKTSN 516
QY 79 P-----QSQALDVNFDOSPIRIGESQSPGLGDMVIEETTPLSLEEL----- 123
DB 517 PELVAQDSEITDYVTITDNLTKYTEEVANMPGLTPDLVOACESELNEVTGKIAVETK 576
QY 124 -----FAQES-----TEMGINPNDYIPEYQGEOP-NSEV-----VVPPTL 157
DB 577 MDLVOTSEVMQESLYPAQOLCPSEFESEATPSPVLPDIYMEAPLNSAVPSAGASYIOPSS 636
QY 158 EPEKGLIKRLYARLFNDGVNKPRLKAKFYOSSQSGETSAIGSSHQKTEPYANIKAALE 217
DB 637 SPLKSSVN--YESIKHEPNRP-----YEEAMSVSLKTKYSGIKELKEPENINAAQ 688
QY 218 DITQ---ESAMDLNLSIRLROTALVAARAVGYDI-DLSITRNSIGEVYIITHL---G 270
DB 689 ETEAPYISICDL-----IKETKLSAEPAPDFSDYSEMAKVEQVPVPHSELVEDSSPDS 742
QY 271 EPPYIDYRAVEVRGEGADDKAFTV---ADEVPLDIGVFHNGKYEY-----KKNLIEVA 322
DB 743 EPPVLD-----FSDDSIPDVPOKODEYVMLVKESLTETSESMLEYENKELSLA 791
QY 323 SAHNG--YFDG--RWLDRSVDIPLPNTADVSLIYDTGTQYRFDEVVEFTTIDPKTNQLT 378
DB 792 PPEGKPYLESFKLSLDNTKDTLLPDEVSTLS----- 823
QY 379 DDDKLPVKKELLEQLLYVMGAEYVNLQAVRALSNDLIATRYENMVTETVFPFERQIQND 438
DB 824 KKEKIPLO---WEELSTA-----VYSND-----DLFISKEAQIRRT 856
QY 439 QVSFEQSSSRTEPAQVDESTLEPIETVELTDGILMDISPIEFSASNLIQDKLNLVAAK 498
DB 857 E-TFSDSS-----PLEITDEPFTLISSTKTDSPSKLAREYTDLEVSH 896
QY 499 ARHLVMPDDR---VLAINHDDGVNRSILGRISDAV-----SAVARAIL--PD- 541
DB 897 KSEIANAPDAGSLPCTELPHDLST-KNIQPKVEEKISFSDFSKNGSATSKVLLLPDV 955
QY 542 -----ESENEVIDLPE-RTALANKRTPADYQSKVPLVYFA 578
DB 956 SALATQAEIESIVKPKVLVKEAEKLLPSDTEKEDRSPSAIFSA 998
```

```
RESULT 40
US-09-893-348-23
Sequence 23, Application US/09893348
Patent No. US20020072493A1
GENERAL INFORMATION:
```

```
APPLICANT: EISENBACH-SCHWARTZ, Michel
APPLICANT: COHEN, Irun R.
APPLICANT: BESERMAN, Pierre
APPLICANT: MOSONOGO, Alon
APPLICANT: MOALEM, Gila
TITLE OF INVENTION: ACTIVATED T-CELLS, NERVOUS SYSTEM-SPECIFIC ANTIGENS AND THEIR
FILE REFERENCE: EIS-SCHWARTZ-2A
CURRENT APPLICATION NUMBER: US/09/893,348
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 09/314,161
PRIOR FILING DATE: 1999-05-19
PRIOR APPLICATION NUMBER: US 09/218,277
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: PCT/US98/14715
PRIOR FILING DATE: 1998-07-21
PRIOR APPLICATION NUMBER: IL 124500
PRIOR FILING DATE: 1998-05-19
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn version 3.1
SEQ ID NO 23
LENGTH: 1192
TYPE: PRT
ORGANISM: Homo sapiens
US-09-893-348-23
```

```
Query Match 2.5%; Score 119.5; DB 10; Length 1192;
Best Local Similarity 18.0%; Pred. No. 6.1;
Matches 116; Conservative 99; Mismatches 241; Indels 187; Gaps 29;
```

```
QY 19 AYLPMTSQLAQQNNPANIINHVAHDIAINQAKAGNPVLLTPQIOARLNAAGLNAAK 78
DB 460 AYTCAPFNPAATESIATNIF---PLGDPITSEKNTDEKKIEKKAQIYTEKNTSTKTSN 516
QY 79 P-----QSQALDVNFDOSPIRIGESQSPGLGDMVIEETTPLSLEEL----- 123
DB 517 PELVAQDSEITDYVTITDNLTKYTEEVANMPGLTPDLVOACESELNEVTGKIAVETK 576
QY 124 -----FAQES-----TEMGINPNDYIPEYQGEOP-NSEV-----VVPPTL 157
DB 577 MDLVOTSEVMQESLYPAQOLCPSEFESEATPSPVLPDIYMEAPLNSAVPSAGASYIOPSS 636
QY 158 EPEKGLIKRLYARLFNDGVNKPRLKAKFYOSSQSGETSAIGSSHQKTEPYANIKAALE 217
DB 637 SPLKSSVN--YESIKHEPNRP-----YEEAMSVSLKTKYSGIKELKEPENINAAQ 688
QY 218 DITQ---ESAMDLNLSIRLROTALVAARAVGYDI-DLSITRNSIGEVYIITHL---G 270
DB 689 ETEAPYISICDL-----IKETKLSAEPAPDFSDYSEMAKVEQVPVPHSELVEDSSPDS 742
QY 271 EPPYIDYRAVEVRGEGADDKAFTV---ADEVPLDIGVFHNGKYEY-----KKNLIEVA 322
DB 743 EPPVLD-----FSDDSIPDVPOKODEYVMLVKESLTETSESMLEYENKELSLA 791
QY 323 SAHNG--YFDG--RWLDRSVDIPLPNTADVSLIYDTGTQYRFDEVVEFTTIDPKTNQLT 378
DB 792 PPEGKPYLESFKLSLDNTKDTLLPDEVSTLS----- 823
QY 379 DDDKLPVKKELLEQLLYVMGAEYVNLQAVRALSNDLIATRYENMVTETVFPFERQIQND 438
DB 824 KKEKIPLO---WEELSTA-----VYSND-----DLFISKEAQIRRT 856
QY 439 QVSFEQSSSRTEPAQVDESTLEPIETVELTDGILMDISPIEFSASNLIQDKLNLVAAK 498
DB 857 E-TFSDSS-----PLEITDEPFTLISSTKTDSPSKLAREYTDLEVSH 896
QY 499 ARHLVMPDDR---VLAINHDDGVNRSILGRISDAV-----SAVARAIL--PD- 541
DB 897 KSEIANAPDAGSLPCTELPHDLST-KNIQPKVEEKISFSDFSKNGSATSKVLLLPDV 955
QY 542 -----ESENEVIDLPE-RTALANKRTPADYQSKVPLVYFA 578
DB 956 SALATQAEIESIVKPKVLVKEAEKLLPSDTEKEDRSPSAIFSA 998
```

Tue Apr 29 12:22:31 2003

us-09-914-168-2.rapb

Page 29

Search completed: April 28, 2003, 16:32:38  
Job time : 97 secs

---



GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:19:19 ; Search time 51 Seconds  
(without alignments)  
1732.305 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 4727  
Sequence: 1 MSKPYLFANRSEMPVALAAY.....TGVEKNPDKLHFFIGRPF 919

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	615.5	13.0	579	2 F83327	conserved hypothet
2	494	10.5	578	2 AE0428	probable exported
3	479	10.1	577	2 AD1055	probable exported
4	454	9.6	582	2 F82064	conserved hypothet
5	436.5	9.2	578	2 B64012	hypothetical prote
6	433.5	9.2	577	2 S56445	hypothetical 64.8K
7	433.5	9.2	577	2 F91278	hypothetical prote
8	433.5	9.2	577	2 F86119	hypothetical prote
9	347	7.3	617	2 G87206	conserved hypothet
10	259	5.5	797	2 D82000	outer membrane pro
11	259	5.5	797	2 AE8128	outer membrane pro
12	236.5	5.0	617	2 AE2897	conserved hypothet
13	236.5	5.0	641	2 H97672	hypothetical prote
14	232.5	4.9	774	2 D97527	group 1 outer memb
15	232.5	4.9	774	2 AE2746	hypothetical prote
16	225	4.8	861	2 S77409	hypothetical prote
17	223.5	4.7	781	2 AH3355	outer membrane pro
18	220	4.7	615	2 AE82025	outer membrane pro
19	216.5	4.6	739	2 AB1430	probable outer mem
20	216	4.6	628	2 B87448	outer membrane pro
21	215	4.5	635	2 G81003	conserved hypothet
22	214	4.5	784	2 E82731	outer membrane ant
23	211.5	4.5	795	2 AC0129	conserved hypothet
24	209	4.4	623	2 A13488	probable surface a
25	206.5	4.4	916	2 G64601	outer membrane pro
26	205.5	4.3	768	2 B97725	outer membrane pro
27	202.5	4.3	768	2 B97725	outer membrane pro
28	197.5	4.2	833	2 AE2089	hypothetical prote
29	194.5	4.1	906	2 F71910	probable outer mem

30	193.5	4.1	676	2 AE2417	hypothetical prote
31	190.5	4.0	769	2 F87486	outer membrane pro
32	183.5	3.9	797	2 H83190	probable outer mem
33	178	3.8	810	2 A64742	hypothetical prote
34	178	3.8	810	2 C90651	hypothetical prote
35	178	3.8	810	2 C85502	hypothetical prote
36	177.5	3.8	778	2 C70412	outer membrane pro
37	177.5	3.8	803	2 B82099	surface antigen VC
38	171.5	3.6	797	2 JC4078	protective surface
39	171	3.6	853	2 D70304	hypothetical prote
40	170.5	3.6	588	2 AD2445	hypothetical prote
41	168	3.6	808	2 F64102	protective surface
42	159.5	3.4	1319	1 C43735	bcsC protein - Ace
43	159	3.4	475	2 T11586	hypothetical prote
44	152.5	3.2	491	2 AC1816	hypothetical prote
45	149.5	3.2	803	2 AB0530	outer membrane pro

ALIGNMENTS

RESULT 1	F83327	conserved hypothetical protein PA2543 [imported] - Pseudomonas aeruginosa (strain PAO C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000	#sequence: revision 15-Sep-2000	#text: change 31-Dec-2000
C:Accession: F83327	R:Stover, C.K.;	Pham, X.Q.; Ewlin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardb, K.; L	.; Lory, S.; Olson, M.V.	
Nature 406, 959-964, 2000	A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa	
A:Reference number: AB29550; M01D:20437337; PMID:10984043	A:Accession: F83327	
A:Status: preliminary	A:Molecule type: DNA	
A:Residues: 1-579	<STO>	
A:Cross-references: GB:AE004682; GB:AE004091; MID:g9948598; PIDN:AA05931.1; GSPDB:GN	A:Experimental source: strain PA01	
C:Genetics:	A:Gene: PA2543	
Query Match	13.0%;	Score 615.5; DB 2; Length 579;
Best Local Similarity	24.5%;	Pred. No. 1.8e-27;
Matches 178; Conservative 117; Mismatches 220; Indels 213; Gaps 17;		
QY 210 ANIKALIEDITQPSAMDLSIPLROTALV---AARAVGYDIDLSIRNSIGEDVLI 265		
Db 41 ANIEAVVGSIGERD---EALQRFRRNMEAQAQKAAQALGTYQ-----AQIDSE 86		
QY 266 IHD-----LGPVYIDYRAVEVNGEGADDKAFTTVADVPVLLIGDVEHHGKYET 314		
Db 87 VKDGKPKKTLTKVPCPEVRLRQVNIQVGEAASLSEFRLPSKQ-QLKPCAKLNGQYED 145		
QY 315 KKKLLINASAEHGYFPGRMIDRSVDVILPDNTADVSLIDTGYQYFDEVEFTTIPKTN 374		
Db 146 AKRLINQASRYGFGQGRSTORSLIDPRAGIADIVDSGGRYTFGKVSF----- 197		
QY 375 QLTTPDKLPVKKLEQLLTVMGSEAVNIQAVRALSNDLIRRYNMVNTLVPPERQ 434		
Db 198 ----DGDST-IEBELLRKRVPRFAGQPYDSELIAELNQNQSSGYF----- 238		
QY 435 IQNDQVSEFOSSSSRTPEAQVDESTLEPIETVELDGIIMDISPIEFSA NLIDKMLN 494		
Db 239 -----EGVRVDAAPYQAQ----- 252		
QY 495 VAAKRAHLVMDMPDRVLAINHDDGVNRSTLGRISDAVSAVARILPDESENEYIDLPERT 554		
Db 253 -----DGAQAI----- 259		
QY 555 ALANRKPADVYQSKKVPVLYFVASDPDQIGLQWGSQDTGRLTKFPHNLINRQGYO 614		
Db 260 -----PVAVRLKARKPTMTGVLGFSITDVGAKARAFNNTTRHVNVAEGHS 302		

```
QY 615 AGAELRLSEDKKVKLYATKPLSHPLNDOLRATLGYOOEFGHSTNGFDLSTRLT---E 670
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 303 LGFSEISAPROWVGAWEYELPLDPPLTDKLRFTSGYOF-----DLVDRESKLLTIGE 356
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 671 HELSRSTIOWGNRRYTSLSKRYLDDKLTQAPPETWODLPDFVNGKRS--QEALLAVAV 728
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 WHSKRP---DGMQRVYSLWMREYKL-----GDSGLSSFLMPDIGY 396
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 729 HKTVADNLVPMGMGROYRSLEYSSGLVSDANMAIARAGISGYVSGDNAYGSRNHRQM 788
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 397 SLETDKVKPDSHCYGRLEQFVWKAKBELLADADVLHYDMAKGLTSPAG-----GRL 449
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 789 TGGIOAGYIWSDFNHNRYRLREFAGDQSIIRGYAHDLSLPSIDKGYLTGGQVLAVGTAE 848
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 450 LGRLOYGIGATNDYKSIPLRFAGGDSQVGRGYDRTLSLSPNSDGDIGGRYMIAGSVE 509
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 849 YNNEFMKDLRLAVFGDIGNVYDKGFTMDTKIGAGVGRMASPVGOVRVDAVATGKEGPN 908
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 510 YQPLAERWRLATFFVDGNAFNSLDPPSIRKTGVGGRVWSPVGPLRLDLAHLDDGCG- 568
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 909 IKLHFFIG 916
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 569 FRLHFSWG 576
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 2

```
AE0428
probable exported protein YPO3524 [imported] - Yersinia pestis (strain C092)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AE0428
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.B.;
demo-Tarrega, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;
H., M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001. PMID:21470413; PMID:11586360
A:Accession: AE0428
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-578 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92753.1; PID:915981448; GSPDB:GN00175
C:Genetics:
A:Gene: YPO3524
```

Query Match 10.5%; Score 494; DB 2; Length 578;

Best Local Similarity 23.4%; Pred. No. 1.6e-20;

Matches 167; Conservative 105; Mismatches 255; Indels 188; Gaps 17;

```
QY 211 NIKAALEDI-TQESAMDLNGSI-PLRLQTLVAVARAYGYD--IDLSIIRNSIGEVYII 266
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 38 NVRAKRLSTIGTDEVYAD--GRFSRVDIAIRQGLRAGLYDPTTFELQNRPARSVYLI 95
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 267 HDL--GEVYIDYRAVEVREGADAKFTTVADVPVLLIGDFPHGKYEYTKNLIENASA 324
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 96 AKVVPGEFVLVAGVDIYLOGGAKTDPDYQALVVRDTPKISSVYNHGFDMFTSSLTGLAL 155
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 325 EHGFYDGRWLDKRSVDYLLPONTADVSLIYDTGYQYRFDEVVFFITIDKTNOLTTDPPKLP 384
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 156 RRYFPANMKISQLGVAQAHEAFWMDIDFDSQGRYRGKVIYF-----OGSQ 201
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 385 VKRELLLEQLLTVMNGEAYNLQAVRALSNDLIATRYENMVMTVEIFEPREQIONQVSEFQ 444
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 202 IREDYIQLNLVPHFEGEYTSDELAELNRRLAATWNSV---VSPD-----FQD 248
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 445 SSSSRTEPAQVDESTLEPIETVELTGTILMDISPIEFSASNLIQDKLNLVAAKARHLYD 504
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 249 AKESK-----LPLD----- 258
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 505 MPDDRVLAINHDDGVNRSILGRISDAVASAARAILLPDESENEVIDLPETALANRKTTPAD 564
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 -----AVTPTRENTV----- 269
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 565 VYOSKAVPLVYFVASDKPRDQIGLGMGSDTGTRLVTKFEHNLINRDQYAGAEIRLSED 624
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 -----ELGGYATVDGPRLTGSMRKPMNNSRCHSLTTTATLAP 308
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 625 KKGYKL-YATKPLSHPLNDOLRATLGYOOEFGHSTNGFDLSTRLTLEHISRSIIONGW 683
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 309 EOTLDESTRIPLEKRNPLEQYLLIOGFFR-----TDLNDNSDTLLNLNARFMDLSGW 362
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 684 NRTYSLRYLDDKLTQAPPETWODLPYDFVNGKPSOEAL--AGVAVHKTVADNLVPMR 741
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 363 QRAINLRMSLD-----HFTGGRVTDITMLLYPCVSIIRTRQRCAMPVW 406
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 742 GYRORYSLENGSSGLVSDANMAIARAGISGYVSGDNAYGSRNHRQHTGGIOAGYISDN 801
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 407 GDSORYSIDVSDTWTWSDVDFGIFQAOVWMIRTLGEK-----NRFVARGNVMIETNN 459
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 802 FNHYPYRLRFAGDQSIIRGYAHDLSLPSIDKGYLTGGQVLAVGTAEYNEFMKDLRLAV 861
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 460 FDRVPPSLRFAGDQSIIRGYAHDLSLPSIDKGYLTGGQVLAVGTAEYNEFMKDLRLAV 519
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 862 FGDIGNVYDKGFTMDTKIGAGVGRMASPVGOVRVDAVATGKEGPNIKLHFFIG 916
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 520 FVDSGEAVNNFSKSDLKTGAGVGRMASPVGPIKIDIAPI-GDNETHGYOYFYG 573
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 3

```
AD1055
probable exported protein ytfm [imported] - Salmonella enterica serov. enterica serov
C:Species: Salmonella enterica subsp. enterica serovar typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AD1055
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davies, R.M.; Dowd, L.; White, N.; Farr
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: AD1055
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-577 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD06889.1; PID:916505537; GSPDB:GN00176
C:Genetics:
A:Gene: ytfm
```

Query Match 10.1%; Score 479; DB 2; Length 577;

Best Local Similarity 22.7%; Pred. No. 1.1e-19;

Matches 166; Conservative 104; Mismatches 257; Indels 204; Gaps 17;

```
QY 210 ANIKAALEDITQESAMDLNGSIPRLQTLVAVAR-----ANGYD--IDL 252
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 22 ANVRLKVEGLSEGLEKNVRAQLSTIOSDEVPDRFRFARVDAIREGLKAGYEPTIKF 81
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 253 SIIRNSIGEVYIITHDL--GEVYIDYRAVEVREGADAKFTTVADVPVLLIGDFPHG 310
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 82 DLPPPAKGRVGLARVYTPGQPVLLIGTEVILRGKARKDKYLLALKTRP-AIGTVLNOG 140
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 311 KYETKKNLIENASAEHGFYDGRWLDKRSVDYLLPONTADVSLIYDTGYQYRFDEVVFFITD 370
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 141 DYDNFKKSLTSVSLRKGYDFSEFIKSQGLIALGRHQAQWMDIDYDSEGRYRGVPTF----- 196
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 371 PKTNOLTTDPPKLPYKRELLLEQLLTVMNGEAYNLQAVRALSNDLIATRYENMVMTVEIFP 430
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 197 -----EGSQRREYIQLNLVPHFEGEYTSDELAELNRRLAATWNSV---VAP 243
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 431 EREQIONQVSEFQSSSRTEPAQVDESTLEPIETVELTGTILMDISPIEFSASNLIQD 490
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 244 E-----FEKSRKTIIP----- 255
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 491 KLNVLAAKARHLYDMPDDRVLAINHDDGVNRSILGRISDAVASAARAILLPDESENEVIDL 550
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```



```
Db 256 -----LKGVSPTNTI--- 268
OY 551 PERTALANRKPADVQSKKVPFLVFAVSDKPRDGOIGLGMGSDPTGRLVTKFEHLINR 610
Db 269 -----ETGVGISTDVGRRVAKSMKKPMWNS 293
OY 611 DGYQAGAEIRLSEDKKGVKL-YATKPLSHPLNDLRLATLGOQEVFGHSTNGFDLSTRTL 669
Db 294 YGHSLTSTSTISAPPEQLDFDSYKMKPLKNPLEQYLLVOGGFKRT---DLNDTQDSDSTL 349
OY 670 EHLISNIIIONGGMNRRYSRLRYLRLDKLTQAPPETMODLPEVFNKGPSEBALL-AGVA 727
Db 350 --AVSRWMLSSGMQRIINLRMSFD-----HFTQGAVNTMTMLFYGVGM 391
OY 728 VHKTVADNLVPMRGYRQRYSLSEVGSSGLVSDAMALARAAGISGVSEGDNAVGSNRAHQ 787
Db 392 ISRTBRSGMLPTWGDQSRISVDYSNTAMGSDVDVSLQAKONWIRFLYDR-----HR 444
OY 788 MTGGIAGYIWSDFNHNVPYRLRFPAGGDSIRGYAHSLSPLSDKGYLTGGGYLVAGTA 847
Db 445 FVVRANLGIETGDFDKVPPDLRFPAGGDSIRGYKYSISPKXSDGNLKGASKLATGSL 504
OY 848 EYNYEFMKDLRLAVFGDIGAAYDKGFTNDTKIGAGVGRMASPVGOVAVD-VATGVKEE 905
Db 505 EYQYNTGKMGAVFVDSGEAVSDIRSRDFKTGTGVGRMASPVGKLDFAVPGDKDE 564
OY 906 GNPRIKLHFFIG 916
Db 565 HG---LQFTIG 572

RESULT 4
F82064
conserved hypothetical protein VC2548 [Imported] - Vibrio cholerae (strain N16961 serog
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82064
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoli, I.; Sellers,
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; M0ID:20406833; PMID:10952301
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-582 <HEI>
A:Cross-references: GB:AE004323; GB:AE003852; NID:99657119; PIDN:AAF95689.1; GSPDB:GN001
A:Experimental source: serogroup O1, strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2548
A:Map position: 1

Query Match 9.6%; Score 454; DB 2; Length 582;
Best Local Similarity 22.4%; Pred. No. 3e-18;
Matches 164; Conservative 104; Mismatches 237; Indels 228; Gaps 20;

OY 211 NIKAALEDITQESAMLNLSI---PRLROTALVAARAVGY--DIDLSTIRNS----- 258
Db 48 NVEAIISSI--AAOYSTSLRPOQSLERSMTBALNALGYHPSIDFTVSEDNQRLRAAV 104
OY 259 -----IGEVDTIHDLEPVIYIDYRAVEVREGADKAFTVADEVPLLDIVFHGK 311
Db 105 TLGEVTRLESEVDIYI-----KGEAGDRDFORLIRRSGRVAPLPHSL 148
OY 312 YETKKNLINASAENHGYFDGRWMLDRSVYILPNTADVSLTYDTGYRDEVEVFFTIDP 371
Db 149 YDNLSKGINRLALQKGYFNGDFQASRLLEVLPENQARVILHPSGIRYLE----- 198
OY 372 KTNQULTDPPDKLPVKRELLQOLLTVNMGAEAYNIQAVRALSNDLIATRYENMVTEIYPE 431
Db 199 --GATTVESQIDENKVM--SLRPFKQGEPIYLSQVGEFNQNLNSNTDMSVVE---PD 251
```

```
OY 432 REQIIONDVSEQSSSSRTEPAQVDESTLEBPVETVELTDGILMDISPIERSASNLQDK 491
Db 252 LSQI----- 255
OY 492 LNLVAAKARHLIYMPDDBRVLAINHDDGVNRSILGRISDAVSAVARAILPDESENEVIDLP 551
Db 256 -----DEG----- 258
OY 552 ERTALANRKPADVQSKKVPFLVFAVSDKPRDGOIGLGMGSDPTGRLVTKFEHLINR 611
Db 259 -----RELPIKVTLLAPQARNOLETGIGYSTDVGVRSLMKKPMWNSQ 301
OY 612 GYQAGAEIRLSEDKKGVKL-YATKPLSHPLNDLRLATLGOQEVFGHSTNGFDLSTRTL 670
Db 302 GHSDDSSSLSPISQITTAGIKIPLDALNRYIYQYGMK-----LDKRDTESE 352
OY 671 HEIS--RSIIIONGGMNRRYSRLRYLRLDKLTQAPPETMODLPEVFNKGPSEBALLAGVA 728
Db 353 SNLSLERHWQLDGGMHRTVFIRYLLENYR-----QGLQDD-----NSQFILPGMTY 398
OY 729 HKT-VADNLVPMRGYRQRYSLSEVGSSGLVSDAMALARAAGISGVSEGDNAVGSNRAHQ 787
Db 399 TRTTRRSNGLLTWGDKQTITLLEYGDPALLSETRYLRLOTGSSWMLRYARN-----HR 451
OY 788 MTGGIAGYIWSDFNHNVPYRLRFPAGGDSIRGYAHSLSPLSDKGYLTGGGYLVAGTA 847
Db 452 ALVAVDGGANLVDEFDQSLRFPAGGDNLRIGYKYSIPQDASGLTAKTYATISSI 511
OY 848 EYNYEFMKDLRLAVFGDIGAAYDKGFTNDT--RIGAGVGRMASPVGOVAVDVAATGV-K 903
Db 512 EYQYRLTGNNMAAFMNVGDAF-----NDNPEMKKGVTGTGRWISPVGRILDLRAWGLDA 566
OY 904 EECNPRIKLHFFIG 916
Db 567 APGDEFKIHFTIG 579

RESULT 5
B64012
hypothetical protein H10698 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: B64012
R:Reischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirsch, E.F.; Kerlavage
, Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; M0ID:95350630; PMID:7542800
A:Accession: B64012
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-578 <TRGR>
A:Cross-references: GB:U32752; GB:U42023; NID:91573692; PIDN:AAC22357.1; PID:91573700

Query Match 9.2%; Score 436.5; DB 2; Length 578;
Best Local Similarity 20.5%; Pred. No. 3e-17;
Matches 142; Conservative 105; Mismatches 235; Indels 211; Gaps 14;

OY 243 RANGYVIDSIIIRNS-IGEVDTIHDL--GEPIYIDYRAVEVREGADKAFTVADEV 299
Db 73 RVEGYESSVRFERKORQGRDLIAHTVPEGPKIAGTVOJEGEAADENFALKNL 132
OY 300 PLLIGDVFHGKYEYTKKNLJENASAENHGYFDGRWMLDRSVYILPNTADVSLTYDTGYR 359
Db 133 P-KDGYLVEHQYTDYKTAISRLALNNGYFDGNFKISRLSEPTHQAMRMLEDSGVRY 191
OY 360 RDEVEVFFTIDPPTNDLTPDKLPVKRELLQOLLTVNMGAEAYNIQAVRALSNDLIATRY 419
Db 192 HYNITF-----SHSQIRDDYLVNINLIKSSDDPYLMNLSLDTSDPFSNM 237
```

```
QY 420 FNNMNTVEIVEREIOIONDOVSFEQSSSRTEPAQVDESTLEPIVETVELTDGILMDSP 479
    ||:|
Db 238 FSSY----- 241
QY 480 IEFASNLIDQDKLNLVAAKARHLYDMPDDRVLAINHDDGVNRSILGRISDAVASARAIL 539
    ||:|
Db 242 -----LVQPNV----- 248
QY 540 PDESENEVIDLPERTALANRKTTPADYQSKKVPILYEVASDKPRDGOIGLQWSDPTGTRL 599
    ||:|
Db 249 -----HKSKTVDVEIILLPRKKNAMELVGVFSTGCVHG 282
QY 600 VTKEHNLINRDYQAGAEHLRSEDKKGVK-LYATKPLSHPLNDQLRATLGYQGEVEGHS 658
    ||:|
Db 283 QIGTKFWINSRGLSLRSLNLTLSAPKOTLEATYKMPKLKNLYYDFAVGME---GEK 338
QY 659 TNGFDLSTRLLEHISRSIIIONGMNRTYSLRYRLDKLQAPPEMODLPEVFNKPS 718
    ||:|
Db 339 EN-DTMTRYLTLSALRYWNNAHGMOYFGGLRMRYDSF-TQAD----- 378
QY 719 QEALLAGVAVHKIYADNLVNPNGYR-----QRYSLVEGSSGLVSDANMA 763
    ||:|
Db 379 -----ITDKTL---LYPTVGFTRTRLRGSPATWDYQKITFDLSKRIMWSESEFI 427
QY 764 IARAGISGVYSGDNAYGNSNRAHOMTGIOAGYIMSDNFNHPYRLRFPFAGDQSRGYA 823
    ||:|
Db 428 KVOASAMVRYTAEN-----HRVVARAEIGYLTHTKIEKIPTRLRFAGDQSRVRYG 480
QY 824 HDLSLPSIDKGYLTGOVLAVGTAEVNEEFMKDLRLAVFGDIGNAVYDKGTNDTKIGAV 883
    ||:|
Db 481 YKKIAPKNRNCKLVGSGRLTLTTSLEYQYQVYPMWMAATFADSGLAADNYATKELRYGTGV 540
QY 884 GVRWASPVGQVRVDVATGVKEGSPKILHPIFG 916
    ||:|
Db 541 GVRWASPVGATKEDIAATPRDKNSKNIQFYIG 573
```

RESULT 6  
S56445  
hypothetical 64.8K protein (msra-chnb1 intergenic region) - Escherichia coli (strain K-12)

N:Alternate names: hypothetical protein o577  
C:Species: Escherichia coli  
C>Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 01-Mar-2002  
C:Accession: S56445; G65233  
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
Nucleic Acids Res. 23: 2105-2119, 1995  
A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.  
A:Reference number: S56314; MUID:9534362; PMID:7610040  
A:Accession: S56445  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-577 <NR>  
A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97116.1; PID:g537061  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: G65233  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-577 <BLAT>  
A:Cross-references: GB:AE000493; GB:U00096; NID:g2367360; PIDN:AAC71177.1; PID:g1790666;  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: ynfM  
A:Start codon: GTG

Query Match 9.2%; Score 433.5; DB 2; Length 577;  
Best Local Similarity 22.9%; Pred. No. 4.4e-17;  
Matches 164; Conservative 100; Mismatches 262; Indels 189; Gaps 18;

```
QY 211 NIKAALEDITQESAMDLNGSIPRLQRTALVAAVAGY-----DIDLSTIIRNSIGVDYII 266
    ||:|
Db 38 NVRAQLSTIESDEVTPPRRFRARVDAIREGLKALGYOPTIEFLRPPPKRGQVLLAK 97
QY 267 HDLGEPPVITYRAVEVRGEGADKAFYTVADVPILLIDVFHHGKETKKNLIENASAEH 326
    ||:|
Db 98 VTPGVPLVIGGTDVLRGARGDRDKYLTLDTRP-AIGVLNQGQYENPKSLSTIALRK 156
QY 327 GYFDGRWIDRSVDYLLPNTADVSLIYDTGTOYRDEDEVVFFIIDTKTQTLTDPKLPYK 386
    ||:|
Db 157 GYFDESEFTKADLGALGLHKAFWMDIDYNSGERYRGHVTF-----EGSQIR 202
QY 367 RELLEQLLTVMNGEAYNLQAVRALSNDLITATRFYMMVMTVEIVEREIOIONDOVSFEQSS 446
    ||:|
Db 203 DEYLDNLVPEKGGDYESKDIAELNRRLSATGEMFVS---VYAPQ-----FDKAR 249
QY 447 SSRTEPAQVDESTLEPIVETVELTDGILMDSPIEFSASNLIDQKLNLAAKARHLYDMP 506
    ||:|
Db 250 ETKVLP-----LT-GV---VSP----- 262
QY 507 DDRVLAINHDDGVNRSILGRISDAVASARAILPDESENEVIDLPERTALANRKTTPADY 566
    ||:|
Db 263 -----RTENTT----- 268
QY 567 QSKKPLYVYFASDKPRDGOIGLQWSDPTGTRLVYKFEHNLINRDYQAGAEHLRSEDKK 626
    ||:|
Db 269 -----ETGVGYSTDVGPRVYKATWKKPMNNSGSHLTSTISAPBQ 309
QY 627 GVKL-YATKPLSHPLNDQLRATLGYQGEVFGHSTNGFPLSTTLLEHISRSIIIONGMNR 685
    ||:|
Db 310 TLDPSYKMKPLLNPLEQYLYVGGFKRT---DLMDTSDSTTL--VASRYMDLSSGQR 363
QY 686 TYSRLRYRLDKLKTQAPPEFWODLPYDFVNGKPSQALN--AGVAAHKIYADNLVNPNGY 743
    ||:|
Db 364 AINLRWSLD-----HFTQGETTNTMLPYEGVMISKRISRGGLPMTGCD 407
QY 744 RORYSLEVGSSGLVSDANMAIARAGISGVYSGDNAYGNSNRAHOMTGIOAGYIMSDNFN 803
    ||:|
Db 408 SQRSIDYSNAMGSDVDVPSVQAOQNMWIRTLYR-----HRVYTRGTGLGMIETGDFD 460
QY 804 HVPYRLRFPFAGDQSRIRGYAHDLSLPSIDKGYLTGOVLAVGTAEVNEEFMKDLRLAVFG 863
    ||:|
Db 461 KVPPLRFPFAGDQSRIRKYKYSIAPKANGDLGASKRLIGSLSYQYVNTGKMWGAFFV 520
QY 864 DIGNAVYDKGTNDTKIGAGVGRWASPVGQVRVDVATGV--KEENPILKILHPIFG 916
    ||:|
Db 521 DSGEAVSDIRSRSDFTGTGTGVGRWESPVGPIKLDRAVPYADKDHG---LQFYIG 572
```

RESULT 7  
F91278  
hypothetical protein ECs5198 [imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: F91278  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8: 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: F91278  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-577 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA938621.1; PID:g13364675; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECs5198

Query Match 9.2%; Score 433.5; DB 2; Length 577;  
Best Local Similarity 22.9%; Pred. No. 4.4e-17;  
Matches 164; Conservative 100; Mismatches 262; Indels 189; Gaps 18;







```
QY 464 IETVELTDCIIMDISPIEFSASNLIQDKLNILVAARHLIYMPDDRLVLAINHDCV---- 519
| : : : |
Db 284 --TIEVSEG-----KHRYFGVAGQY-----STTEGIGLOG 311

QY 520 ---NRSLIG-----RISDAVASARAILPDESENEVIDLPE-----RTA 555
| : : : |
Db 312 YMGHNLFGQASLRIEBSVSRIAE-----SSVEGMDYSAGITFTKPGMFNPTFTKTS 366

QY 556 LANRRTPADVYQSKVPLVYFVASDKPRDQIGLGWSDTGRILYTKFEHNLINRDGYOA 615
| : : : |
Db 367 LIATKENDPTVRAKTL-----TGT---AGFAYELNDTDTAA 400

QY 616 GAELRL--SEDKKGVKLYATKPLSHPL-----NDQLRATIGYQOEFGHSTNGFDLST 666
| : : : |
Db 401 GLEVOAMDTEDAFGKNEYLT--TSIPLEFVRDTRDKLNPTEGFRAS----- 445

QY 667 RTLEHEISRSIIQNGWNRRTSLRYRLDKLKTQAPPETWQDLPVDFVNGKPSQEALLAGV 726
| : : : |
Db 446 -----LAKPSYEAL----- 455

QY 727 AVHKTVDNLVNPARGYRQRYTSLEVSSGLVSDANMAIARAGISGVYSGDNAYGSNRH 786
| : : : |
Db 456 --NGTFSSFECSITGYK-----GLGAEDRLIM-----AG 483

QY 787 QMTGIGAGYIWSDNFNHVRPLRFPAAGDOSIRGVAHDSLPSIDKGYLTGGVLAAGT 846
| : : : |
Db 484 KLSGGVLVG--GSDLDIPITRRFRFAGGGSVGRYSVOEISPYNAAGADATGGRSVYGS 540

QY 847 AEYNEFMKDLRLAVFGDIGNAYDKGFT---NDTRIGAGVGRMASPVQGVVDVATGVK 903
| : : : |
Db 541 VEARIKVTDTIGLVPPFDAGVSD-GVTPDFSDIRAGAGIGLRATPGRPLRLDVAMPLE 599

QY 904 --EEGNPKIKLHFITGTPF 919
| : : : |
Db 600 KYDGGNFGIYAGIGQSF 617

RESULT 13
H97672
-hypothetical protein AGR_C_4742 [Imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: H97672
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: H97672
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-641 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK8337.1; PID:g15157817; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_4742
A:Map position: circular chromosome

Query Match 5.0%; Score 236.5; DB 2; Length 641;
Best Local Similarity 20.3%; Pred. No. 9.3e-06;
Matches 162; Conservative 107; Mismatches 246; Indels 283; Gaps 31;

QY 200 GSHQKTEPANKALEDTTQESAMLNSTI-PRLKOTALVARAVGYDIDILSIIRNS 258
| : : : |
Db 49 GSEPEVEVINPKYA--VTLDA--DADKSLKSSLENSILLADKDKPASGDLGLIKA 104

QY 259 ICEVDYIINDLGEPV-YIDYRAVEVRGEGADKAFITVADE---VPLLI---GDVFNHG 310
| : : : |
Db 105 RODRDLIALIYENARGGIVNVTAGKANVDDLPRNPVDFHSTFVPMVITVTPGPKFTLG 164

QY 311 KYE---TKKNIENASAEHGYFDGRWMLDRSYDVIILPD----- 344
| : : : |
Db 165 NVRLGSDVYGRNIDEXGLIAGDGAGSLAIRAGNKLILDKAGGRPLAKLKKRAVANHA 224
```

```
QY 345 -NTADVSLITDGTQYRFDEVEVFTIDPKTNLTDPDKLPYKRELEOLLTVNNGEAVN 403
| : : : |
Db 225 TMTVTITMAEGGPAPPLGAVTV-----TGETYDGD-----FIRKYSRLNGGEPYS 271

QY 404 LOAVBALNDLIATRYFNWNTIEVPEREQIQNDQVSFOSSSSRTPEAQVDESTLEPV 463
| : : : |
Db 272 PEKLRKAADRILQGVF-----SLLTIKEAGTILAROGTIFL 307

QY 464 IETVELTDCIIMDISPIEFSASNLIQDKLNILVAARHLIYMPDDRLVLAINHDCV---- 519
| : : : |
Db 308 --TIEVSEG-----KHRYFGVAGQY-----STTEGIGLOG 335

QY 520 ---NRSLIG-----RISDAVASARAILPDESENEVIDLPE-----RTA 555
| : : : |
Db 356 YMGHNLFGQASLRIEBSVSRIAE-----SSVEGMDYSAGITFTKPGMFNPTFTKTS 390

QY 556 LANRRTPADVYQSKVPLVYFVASDKPRDQIGLGWSDTGRILYTKFEHNLINRDGYOA 615
| : : : |
Db 391 LIATKENDPTVRAKTL-----TGT---AGFAYELNDTDTAA 424

QY 616 GAELRL--SEDKKGVKLYATKPLSHPL-----NDQLRATIGYQOEFGHSTNGFDLST 666
| : : : |
Db 425 GLEVOAMDTEDAFGKNEYLT--TSIPLEFVRDTRDKLNPTEGFRAS----- 469

QY 667 RTLEHEISRSIIQNGWNRRTSLRYRLDKLKTQAPPETWQDLPVDFVNGKPSQEALLAGV 726
| : : : |
Db 470 -----LAKPSYEAL----- 479

QY 727 AVHKTVDNLVNPARGYRQRYTSLEVSSGLVSDANMAIARAGISGVYSGDNAYGSNRH 786
| : : : |
Db 480 --NGTFSSFECSITGYK-----GLGAEDRLIM-----AG 507

QY 787 QMTGIGAGYIWSDNFNHVRPLRFPAAGDOSIRGVAHDSLPSIDKGYLTGGVLAAGT 846
| : : : |
Db 508 KLSGGVLVG--GSDLDIPITRRFRFAGGGSVGRYSVOEISPYNAAGATGGRSVYGS 564

QY 847 AEYNEFMKDLRLAVFGDIGNAYDKGFT---NDTRIGAGVGRMASPVQGVVDVATGVK 903
| : : : |
Db 565 VEARIKVTDTIGLVPPFDAGVSD-GVTPDFSDIRAGAGIGLRATPGRPLRLDVAMPLE 623

QY 904 --EEGNPKIKLHFITGTPF 919
| : : : |
Db 624 KYDGGNFGIYAGIGQSF 641

RESULT 14
D97527
omp1 protein precursor (U51683) [Imported] - Agrobacterium tumefaciens (strain C58, C
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: D97527
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: D97527
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87173.1; PID:g15156447; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2554
A:Map position: circular chromosome

Query Match 4.9%; Score 232.5; DB 2; Length 774;
Best Local Similarity 21.0%; Pred. No. 2.2e-05;
Matches 178; Conservative 109; Mismatches 310; Indels 251; Gaps 39;

QY 165 IKRLRYARLF-----NDGVNRY---PRLKAFYQSSQSGETSAIGSS 202
| : : : |
Db 77 VKRLIATGIFSNVMSRYSGLTVVTVNENQILVNOYVFNGNKKIKDDKLAGIVQTOPKQPF 136
```

QY 203 HQR--TEPYAIKALLEDITQESANDLNGSIPLRLOTALVAARAGVYIDLSII----- 255  
 Db 137 NQATVTDADIARIKEAYSAGR-----SDVEITQTQV-----SYGGGRVIAAFVINEGE 184  
 QY 256 RNSIGEVDAIIHHDGEPYIDR-AVEYRGEGADKAFTTVADVEPLLIDYFHHGXYET 314  
 Db 185 RTKIGRIDFT-----GNNSYSDGRLAIVNTKSKNMLSPFLTRK-----DVTYNEEDLRA 232  
 QY 315 KKNLIENASAEHGYPFDGRMLDRSVDAIIPD--NTADVSLITDGTQYRFDEVEFEETIDPK 372  
 Db 233 DEEARLQFYRNGYADFRV--SSDAVIDESKNEEYIISTYVDGSKRTYFQGNNAVVESTVP- 289  
 QY 373 TNQTLTPPDKLPVARELLEQLLTVMKEAYNLQAVRALSNDLIATRYFNKNVTEIYEPER 432  
 Db 290 -----GVGSESLQGLVETRGQASYSAREVQQ-SMEALSKR----- 323  
 QY 433 EQIQNDVSPFQSSSSPPEPAVDSTLEPIETETJELTDGILMDISPIERSASNLIDOKL 492  
 Db 324 -----VAGGYPFARATPPGDRDMSCNTIGVTYIVDGERAYVERIETRGTRFRDYV 376  
 QY 493 NLVAAKARHLVDMDDRVLAINHDDGVNRSI-----IGRISDAVASAVARAILPDE 542  
 Db 377 -----IRREED-----ISEGDAPNQITIAAKRRLREALGYFSKVNISTAGGSAPD- 421  
 QY 543 SENEVIDLPERTALANKRTPADPVYOSKAPPLYVFAASDKPRDQIGLGMGSDTGTRLVTK 602  
 Db 422 -----RVYIVVDVEDDOSTGSRFGIAGYSQNDGVILLEAS 454  
 QY 603 FEH-NLINRDGY--QAGAEILRLSEDKGVKLYATKPLSHPLNDQIRATLGYO----- 651  
 Db 455 VEERKNFLGRQYIRVAAG--EDDARTYLSLSTPEPY-----FLGYRLAAGEFL 501  
 QY 652 -----QEVFGHSTNGFEDLSTRLEHETSRISLION-----GWNRTY 687  
 Db 502 FKNQSKSEDDYNNYDEQFAL-----RYTAPITENLSFTFKYTYKQINVEGSGDMQNNNA 554  
 QY 688 SLRYRLDKLKTQAPPEITWQDLPVDFVNGKRPSEQALLAGYAVHRYTADNLVNPERRGTORY 747  
 Db 555 NL-----AEPQAL-----IRGEDWTQSLSTNLTANNTLIDDRNM-PREGQOAL 597  
 QY 748 SLEVGSSGLVSDANMALARAGISGVSPGD--NAVGSNRHOMTGTGIOAGYISDNPNHV 805  
 Db 598 TNEF--AGLGSDSYXXIYLAARYITLSDIEDYDGS-----LTG--QAGHWPRTGDNL 648  
 QY 806 PYLRFPFAGGDSQIRGYAHSLSPT--SDKGYLTGGOVLAVGTAEXN-----YEFMKD 856  
 Db 649 VFD-QEKFQGRQ--VRGEKNDIGIPRIGSDS--IGTTYFAASAEVTAMPQVDFP--G 701  
 QY 857 LRLAVFGDIGNAYDKGT-----NDTKIGAGYGVRRASVGVQYRVDAVAGVKEGNP 908  
 Db 702 LRLAGFVDAQTMGNKYSTQTVKDNDSIRASAGIGVMASPFQPIRVYAIPIAKEDYD 761  
 QY 909 IKLHFFIG 916  
 Db 762 EEQRRFRFG 769  
 RESULT 15  
 AE2746  
 group 1 outer membrane protein precursor omp1 [imported] - Agrobacterium tumefaciens (str. C58)  
 C.Species: Agrobacterium tumefaciens  
 C.Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
 C.Accession: AE2746  
 R.Wood, D.Will, S. Stubb, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H.;  
 Erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan,  
 A.; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 A.; Ster, E.W.  
 A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; PMID:11743193  
 A:Accession: AE2746

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-774 <KUR>  
A:Cross-references: GB:AE0086a8; PIDN:AA142387.1; PID:g17739796; GSPDB:GN00186  
A:Experimental source: strain C58 (dupont)  
C:Genetics:  
A:Gene: omp1  
A:Map position: circular chromosome

Query Match	4.9%	Score 232.5	DB 2	Length 774
Best Local Similarity	21.0%	Pred. No. 2.2e-05		
Matches 1/8; Conservative	109	Mismatches 310	Indels 251	Gaps 39

QY 165 IKRLYARLE-----NDGVNKV---PLKAKFYQSSQSGETSAIGSS 202

Db 77 VKRLATGYFSNVSMRVSGSTLVVTVNENQLVNVQVFNNGNRKIKDKLAGIVQTQPMGPF 136

QY 203 HQ--TEPYANIKALEDITQESAMDLSIPLRQOTALVARAVGYDIDLSII----- 255

Db 137 NQAI VTAD IARI KEA YSAIGR-----SDVEIT TQTV-----SVGQGRVNI AFVINEGE 184

QY 256 RNSIGEVDTIHDLGEPVYIDYR-AVEVRGEGADDKAFTTVADEVPLLGDFVHHGKYET 314

Db 185 RTKIGRIDFI-----GNSYSYDGRLLAAVINTKKSMLSFLTR-----DYNNEDKLRA 232

QY 315 KKNLIENASAEHGYFDGRWLD RSV DVI LPD - - NTADVSLIYDTGTQYRDEVEVFTIDPK 372

```

Db      233 DEEALRQFYNRGYADFRV--SSDAVLDESKNEYTISITVDEGKKYDFGNVAVESTVP- 289
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
          ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

```

QY 373 TNOLTPDPDKLPVKRELEOLTVNMGEAYNLOAVRALNDLIATRYENMVNTEIVEPER 432

```
Db      290 -----GVDGSELOGLVETBOGASYSAKEVVO-SMEATSKR----- 323
```

OV 433 EOIONDOVSFEOSSSSRTEPAOVDESTI,EPVTFETVEI,TDGII,MDISPIEFSASNI,IODKI. 492

```
Db      324 -----YAGFGYPARVTPRGDRDMSGNTIGVTYITDQGFPAVEBIFIRGNTEPTNV 376
```

493 NLVAKARHI.YDMPDRVT.ATNHDDGVNRST-----IGPISDAVSAVBPATIBDF 543

```

Db      377      | : | | : | | : | |
          -----TSECDAFNCTTTAAKBPBEAICVFEKVNTSEACCCAND- 431

```

05 543 SENEVITDI.PEPTAI ANBKTBADVYOSKVB I VVEVA SNVBBCOTCI CWCSCMCMBI UMZY 603

pb 423 ----- :| : | | || : | : | |

603 FEB-NI TNPDCV - - - OACAEI DI CEDWCGUET VATEWOT CUNDI WDOI DUEI QWO CCL4

11 : 1 11  
 12 : 1 11  
 13 : 1 11  
 14 : 1 11  
 15 : 1 11  
 16 : 1 11  
 17 : 1 11  
 18 : 1 11  
 19 : 1 11  
 20 : 1 11  
 21 : 1 11  
 22 : 1 11  
 23 : 1 11  
 24 : 1 11  
 25 : 1 11  
 26 : 1 11  
 27 : 1 11  
 28 : 1 11  
 29 : 1 11  
 30 : 1 11  
 31 : 1 11  
 32 : 1 11  
 33 : 1 11  
 34 : 1 11  
 35 : 1 11  
 36 : 1 11  
 37 : 1 11  
 38 : 1 11  
 39 : 1 11  
 40 : 1 11  
 41 : 1 11  
 42 : 1 11  
 43 : 1 11  
 44 : 1 11  
 45 : 1 11  
 46 : 1 11  
 47 : 1 11  
 48 : 1 11  
 49 : 1 11  
 50 : 1 11  
 51 : 1 11  
 52 : 1 11  
 53 : 1 11  
 54 : 1 11  
 55 : 1 11  
 56 : 1 11  
 57 : 1 11  
 58 : 1 11  
 59 : 1 11  
 60 : 1 11  
 61 : 1 11  
 62 : 1 11  
 63 : 1 11  
 64 : 1 11  
 65 : 1 11  
 66 : 1 11  
 67 : 1 11  
 68 : 1 11  
 69 : 1 11  
 70 : 1 11  
 71 : 1 11  
 72 : 1 11  
 73 : 1 11  
 74 : 1 11  
 75 : 1 11  
 76 : 1 11  
 77 : 1 11  
 78 : 1 11  
 79 : 1 11  
 80 : 1 11  
 81 : 1 11  
 82 : 1 11  
 83 : 1 11  
 84 : 1 11  
 85 : 1 11  
 86 : 1 11  
 87 : 1 11  
 88 : 1 11  
 89 : 1 11  
 90 : 1 11  
 91 : 1 11  
 92 : 1 11  
 93 : 1 11  
 94 : 1 11  
 95 : 1 11  
 96 : 1 11  
 97 : 1 11  
 98 : 1 11  
 99 : 1 11  
 100 : 1 11

Q: 653 — ОПРЕДЕЛЕНИЕ СРЕДНЕГО ПОТОКА

[illegible][illegible][illegible]

740 OF BUSINESS AND FINANCE

[illegible][illegible]

000 F I N A N C E S O D Q S I A V G I A H D S U P L -- S D A V G I L I G G Q V E A V G I A E I N ----- Y E F M K D 856  
: : | | | : | | : : | : | : | | : | | : |  
27

070 VLD QALNCOX ANOIRNDGJGFALSDS -- 190111FMASAEVIAFMPOVEDE--G /01

ZY 00 / LKLVFQDIGNMAI DNGFI -----NDIKIGAGVGVRKWASPVGQVRRVDVAJGVKEEGNP 908  
||||| : ||||| :  
||||| : ||||| :

00 / 02 LKLBGF VDAQIMIGNNVSIQIVNDDNSIKASAGIGVMMA\$PFGPIKVDYALPIAKEDYD / 61

QY 509 1KLNFFIG 916  
: 1 1

DD / 02 FEBRUARY / 69

## RESULT 16

S77409

hypothetical protein - *Synechocystis* sp. (strain PCC 6803)  
C:Species: *Synechocystis* sp.

A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C:Accession: S77409

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

S.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S77409

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1861 <KAN>

A:Cross-references: EMBL:D90906; GB:AB001339; NID:g1652492; PIDN:BAJ17512.1; PID:d101824

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Start codon: GTG

Query Match 4.8%; Score 225; DB 2; Length 861;

Best Local Similarity 19.7%; Pred. No. 6,9e-05;

Matches 189; Conservative 130; Mismatches 399; Indels 242; Gaps 38;

93 SPISRIGESPPLGLDMSVIEETPLS-LEELFAOESTEMGINPDYIPEYQGEQPNSEV 151

11 SPYRLLLTSGVLGASPAQATQAFSPMLNLAGDNDTELLVSSSEL-NFPGPTSESEA 69

152 VVPTLEPEKPLIKRLARLFNGVKNKVPRLKAKFYOSSSGESALGSSHQKTEPRAN 211

70 SNVSTLEAPASVINALNGE--NCEISVIP-ELIQIDGDLGGTMEISA-----GN 118

212 IKALEDITQESAM-----LNGSIPRLROTALVAARAVGYDIDLSIRNSIGEDVYI 266

119 LDVGVDLPPAPVDSAEALQANELPHGNNAVAOYAPRAVOEQLIAETKDNDGTHV- 177

267 HDLGEF-----VYIDRAVEYRGEGADKAFYVADEVPLIGV-----FHNG 311

178 ---EPOSPLMAQAAVEEVAEVEATEETTGVTETPEETPPTPAPPTNTEGTRGP 233

312 YETKKNIIENASAEHGFYDGRMLDRSDVILPDMTAD-----VSLYDTGTQYRF 361

234 TQTLPTSTPPAS-----PSTTPAPABEERVLVSEVLVYTTTPEL 274

362 DEVEFTIDPKTQTLTDPDKLPVKRELLEQLLTVMNGEAYNLQAVRALNSDLATFRFN 421

275 ELVY-----NAIRTOPGRTRTRTQLOEDV-----NAIYATGYFS 309

422 MYNTEIYPEREQIQNOVSEFQSSSS-----RTEPAQ-----VDESTLE- 461

310 NVR---VAFSOTPL-GVAVTEYVQANPVFTGLNTRTVPETLEGKRIILLPOEYVDETFEGEQ 365

462 --PIETVELTDLGILMDISPIEFSSANLIQDKLNLVAAKARHLXDMPPDRVLAIHHDGV 519

366 YGKLNLNRELQEG---IKTINEMYSNOGYDLQOVVSP---QVAGDGQVTLVIAGSI 416

520 NRSILGRISDAVSAVARAILPDESENEVIDLPERTALANRK---TPADVYOSKKV----- 571

417 VENIQVRFED-----SEDEPVQGRTRDFIITREMRLLKPGDVFNENRQOTDLQ 463

572 -----PLVYFVASD--KPRGOIGLGWG--SDGTRLVTFEHL 607

464 RYSLGLFEDVRLSFPNGSDPTEYLVNVVYEGNTGTAAGGSSSSGLFCTISYQERN 523

608 INRDYQAGABLRLESDKGVKLYATKPL--SHPLNDQLRATLGYQOEVEFGHSTNGFDLS 665

524 LGGNNQITIGVBAOGRRLFPDVSFTDPMIGDDPRTSYTANL-FRRRTISLVFGADSS 582

666 TRTLEHESKTIQNG-----GNNRTYSLRYLKDCLKTQAPPEWTMDL 708

Db 583 IRTNNGDSPRVVTRGLGTFPRPLADVPAPPMKRLSAGCYQVNRTEANAALSPESA 642

Qy 709 PVDEVNCRP-----SQEALLAGVAHRTVADNLVNMGRYQRYSLIEVSSGLVSDAN 761

Db 643 PLNGFNQPLSFSDYGVDELFTLSFGASQDNRRNNALQPTSGSLVRFCAE--OTIVGTGN 700

Qy 762 MAIRAGISSYVSF-----GDNAVSRBAHQMGIGIAGYIWSDNFNHVPRLRF 811

Db 701 IMTRF--LRGSYSTIYVNMWLDLTFGLVESTIQTPTVAFNVOAGTVLGD---LPTPEAF 754

Qy 812 FAGGDQSIQYAHDSLSPISDKGYLTGGVLAAGTAEVYEFMKDLRLAVEGDGN---- 867

Db 755 ILGGSNSYRGV-----QEGELGNGRSFQATAEKRFPIIAVAGLAFVYDGSNLGS 805

Qy 868 -----AYDKGFLNDRKIGAGVYRNASPPYGOVRVDVATGKRECNPIKHFITGPF 919

Db 806 QCAVPGFPAIVRGILPG-SGVYGLGVRIQSPVPIRIDL--GFTGEGES-RINFIGEXE 861

## RESULT 17

AH3355

outer membrane protein BMEI0830 [imported] - *Brucella melitensis* (strain 16M)  
C:Species: *Brucella melitensis*  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002

C:Accession: AH3355

R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melit*

A:Reference number: AD3252; PMID:11756688

A:Accession: AH3355

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-781 <KUR>

A:Cross-references: GB:AE008917; PIDN:AAL52011.1; PID:g17982775; GSPDB:GN00190

A:Experimental source: strain 16M

A:Gene: BMEI0830

A:Map position: I

Query Match 4.7%; Score 223.5; DB 2; Length 781;

Best Local Similarity 20.9%; Pred. No. 7.1e-05;

Matches 180; Conservative 116; Mismatches 302; Indels 263; Gaps 45;

165 IKRLVA-----RLFNDVNVKVPRLKAK-----FYOSSSGESALGSSHQKTEPRAN 211

Db 78 VKRLFRAMGLFSVRIHQSGSTLIIVQYKERSVNNVLFQGNKKIKDPDLARAVQ-LKPRAP 136

Qy 212 IKALEDITQESAMDLNGSIPRLROTALVAARAVGYDIDLSIRNSIGEDVYIHDLGE 271

137 FDMATMEADKEAIKAYSHIG--RSDATVNAKTV-----DL-----GQGRVNV-VYEINE 183

Qy 272 PYIDYRAVEYRGEGADKAFY--FYADEVP-----LLIGDYFHNGKYEYKKNLIE 320

Db 184 GSRTKIANIEYV---NMAFSGRRLRDVISTKRSNPLSLMTRNDVYDEGRLOADEETLR 239

Qy 321 NASAEHGFYDGRMLDRSDVILPDMTADVSLIYDTGTQYRFDEV-VFTIDPKTQTLTDD 379

Db 240 RYVYNRGADVFLSSNAVLDPSTNEYTLITVYDEGPRYTFGVSVESTYD----- 290

Qy 380 PDKLPVKRELLEQLLTVMNGEAYNLQAVRALNSDLATATFRFNMYNTEIYPEREQIQNO 439

Db 291 ---GVDQALRLVLYKTRGKPSAKEI-----DSVLSTESVAGSG 329

Qy 440 VSEFQSSSRTEPAQYDESLPEVETVELFD-GILMDISPIEFSSANLIQDKLNLVAAK 498

Db 330 YAF-----AKVRP-KDGRFENHTISVYSDGPRAYITRIEIRNDKTRQVY----- 377

Qy 499 ARHLYDMPDRLVLAIHHDGVNRSILGR-----ISDAVSAVARAILPDESENEVI---D 549

Db 378 IRREFD-----LNEGDAFNQVQAKRRLLEVLDFTQVWISYAPSGSEPOVILVYD 429

Qy 550 LPERTALANRKTPADVYOSKKVPLVYFVASDKPRDQIGLGWGSIDG-----TRLVY 601



```
Db 430 VVEKST-----GEFSGIGGYTGGESPGAQVEAAT 460
Oy 602 KEFHNIINRDGY---OAGAEILSEDKKGVKLYATKPL-----SHPLND- 642
Db 461 --ERNLGNQYIRISAGAG---QDDMRNATGLFTEPEYFGLYSAGFDVFRSRVND 515
Oy 643 ----QLRATLGYOOEVEFGHSTNG-----FDLSTRLEHEISRSII---ONGCWN- 685
Db 516 YVEEQTGCTIRFGLPITDNFSAGIAYSLVQEKYDLFRGDAENYAAPRLLEAENSPLRS 575
Oy 686 --TYSLR-YRLDKLTKQAPETWODLPVPFVNKPSOE-ALLAGVA--VHKYVADNLVNP 739
Db 576 SVSYSLYSIDIKN-----PHDGLYKGFIOEFAGLGDGDKKYVTKTEKGN---- 621
Oy 740 MNGYRORYLEVGSSSLVSDANNAIARAGISGYSFSGDNAYGSNRHQMGTGQAGTWS 799
Db 622 ----YYOTLSQEADIVCLL-----VGAGYIHERGDD-----GVRI----- 653
Oy 800 DNEHNPYRLRFAGGDQSLRGYAHDSLSPISD----KGYLTGGQVLAAGTAENYEF-- 853
Db 654 -----FDLFKNSSDIIRGFKNFGIGPYODAKNGKRYMGTTFYFSGTAEVQFPMV 704
Oy 854 ----MKDLRLAVGEGICNATDKGFTNDT-----KIGAGVGV--RWASFVQGVRVAVAT 900
Db 705 LPESLGVKRAFPADAATLYG---NDTPDISGDDKRLRASVGVSLMWASPFGLRFDYAF 760
Oy 901 GV-KEEGNPVK-LHFFIGTFP 919
Db 761 PVAKADTKVQVNFNGVSTKF 781
```

```
RESULT 18
A82025
Probable outer membrane protein NMA0296 [imported] - Neisseria meningitidis (strain 2249
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: A82025
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holtroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: A82025
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-615 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:97378778; PIDN:CA83602.1; PID:9737905
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: NMA0296
```

```
Query Match 4.7%; Score 220; DB 2; Length 615;
Best Local Similarity 22.3%; Pred. No. 7.6e-05;
Matches 131; Conservative 72; Mismatches 236; Indels 148; Gaps 19;
```

```
Oy 462 PVIETVELTDLGILMDISPIEFASNLIDKLVAKARHLCLDMDDRLAINHDDGVNR 521
Db 48 POTESVTKLKPFPVRIODSEIKDMVEHPLITQOOEEVLDKQGTGFLAEADPNVKT 107
Oy 522 SI--LGRISDAVSA-----VARAILPESENEVIDLPERTALAN 558
Db 108 MLRSKGYFSSKYSLEKDGAYVHTTGPRTKRIANVVAIIIDLSDCGLAEYYNALEN 167
Oy 559 RKTPA-----DYQSK-----VPLVEVASDKP 582
Db 168 WQDPVSGSDFDODSMNSKTSYLGAVTRKAYPLAKLGNTRAAVNPDTATADLNVVVDGSRP 227
Oy 583 RDQIGLQWGSDTGTR-----LVTKFEHNLINRCGYQAGEL- 619
Db 228 ----IAGDFEITGTQKRPQOIVSGIARFQETPPDLDLDDQAL-EQNGHSGASVQ 282
Oy 620 ----RLSEDKKGVKLYATKPLSHPLNDQLRAT---LG-----YOOEVEFGHSTNG 661
```

```
Db 283 ADFRRLQGDRAVPYKVSUTEVKRHKLENGIRLDSEYGLGKRIAYDYNLKFNGYIGSVYMD 342
Oy 662 FDLSTRLEHEISRSITQNGW---NRTYSLRRLDKLTKQAPETWODLPVDFVNGKPS 718
Db 343 MDXYETTLAAGISOPRNYRGNMYTNSVNSRSTQNLEKRAFSGGIYVDRAGIDARLG 402
Oy 719 QEALLAGVAV-----HKT-----VADNLVNMRCGRQYSLSEVGSSGLVSD 759
Db 403 AEFLAECRKLPISDIDLGNSHATMLTASMKRQQLNNVLNGBENGHYLDGKTGLGAEFLS 462
Oy 760 ANN--AARAGISGYVSFGDNAYGSNRHQMGTGI---OAGY-IWSDNPNHYPRLRF 812
Db 463 TALIRTSARAG-----YFFPENKKGITGFIIRGAGATVADNAN-VPSGLMFR 510
Oy 813 AGGDQSLRGYAHDSLSPISDKGYLTGGQVLAAGTAENYEFMDRLAVFGDIGNAYDKG 872
Db 511 SGGASSVRYGYELDSIGLAGPNSVLPERRALLVGSLEYQLPFTRTLSCGAVFHDMDAANF 570
Oy 873 FTNDTKIGAGVGRMASPVQGVAVDVAATGVKEGPNFKLHFFIGTFP 919
Db 571 KRMKLHSGSLGVWFSPILAPFSFDIAYGHSDK--KIRWHISLQTRP 615
```

```
RESULT 19
A81430
Outer membrane protein Cj0129c [imported] - Campylobacter jejuni (strain NCCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: A81430
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chli
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; Vanlille, A.; Whitehead, S.; Bar
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: A81430
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-739 <PAR>
A:Cross-references: GB:AL139074; GB:AL111168; NID:96967505; PIDN:CA872613.1; PID:9696
A:Experimental source: serotype O2, strain NCCTC 11168
C:Genetics:
A:Gene: Cj0129c
```

```
Query Match 4.6%; Score 216.5; DB 2; Length 739;
Best Local Similarity 18.1%; Pred. No. 0.00016;
Matches 172; Conservative 142; Mismatches 317; Indels 317; Gaps 43;
```

```
Oy 68 ARLNMA-----GLNAKPGQALDVNFDQDSPSRIGEGSPPLGLDMSVLEETTPIS 119
Db 13 AMANNAATTIDIKFICGILNLSNTSAINIAGL-----KIGEEINPAKIMVAILN---LY 61
Oy 120 LEELEFAOESTENGINPNDIYPEYQEQPN-SEVVVPTPLEKPGF-----TKRLVRLF 173
Db 62 KKNYFENIATIVE---NNNGLEIITVETKPIYATYTTIGCINSNRKQVESLIGLKR--GTL 116
Oy 174 NDGVNK--VPRKANFYQSS-----OSGETSAIGSSH--- 203
Db 117 DEGNIKAEIERIKAYEAKSYFDYVEYKKTKLENTDGLLEPTIVRGNGIITIDNVHLSG 176
Oy 204 OKTEPYANKALLEDITQP-----SANDLNSITRLKOTPAIVAAANGYDI 250
Db 177 AKKESYSIDIEPAVNVKKEFEKMMGMGRNDGKLKVELSLSDSSRIADEVM---KKGYL 232
Oy 251 DLS-----IIRNSIGEVDIITHDGEPEYIDYRAVEVREGADAKFAFTVADEVPLIG 304
Db 223 QVSSPYLKTYYDTTYQANLTYFIKE-GKPKT--KSIISLENPLFDKQANQATYKDLRSSAG 289
Oy 305 DVFHGKYETKKNLIENASAEHGYFDGRMLDRSDVDVILPDNTADVSLIYDTGTQYRFDEV 364
Db 290 KTIINIEDIRKDVKTETQESADLGY---AFVEEYPDIOKNDQGEATVYFKV---IPHKV 343
Oy 365 VFTTIDPTNQLTTPDPDLKPVKRELLEQLLYNMEAYINLQAVRLASNDLIATRTFNMVN 424
```

```

Db 344 YIRNIIISGSRIVD---RVYIRRELY-----ITEGNLYNRTDISESKNALKRTSYFDDVN 395
QY 425 TELVPERREQIQNDQVSEFGSSSRTEPAQVDSSTLEPVEIETEL-----TD 471
Db 396 -----IKEEKVDDTHIDLIVDKFAKSAISGIGYSSD 430
QY 472 GILMDISPIEFASNLQDKLNLVAAKARHLVDMDDRVLAINHDDGVNBSITGRISDAV 531
Db 431 GILLNAS---LSPDNIIFGSGI-----KSSVSVKSSD-----TLSGRIS--- 466
QY 532 SAVARALPDESENEVIDLEPRTALANRKTTPADYQSKKPLVYFVASDKPR--DQIGL 589
Db 467 -----LVN-----PRVLDQSYSL 479
QY 590 GMSGDTGRLVTKFEHNLINRDGYQAGAEIRLSEDKKGVLYATKPLSHPLNDQLRATLG 649
Db 480 G-----GT-----LYS-----ND----- 487
QY 650 YQGEVGHSTNGFDLSTRLEHEISRTIIONGKNRTYSLRYRLDKLTKQAPPETWODLP 709
Db 488 YEMDNVSEKNYGFDI---TIGROFARY-----NVSLTYNLEQ----- 522
QY 710 VDPVNGCPSEALLAGVAVKTVADNLVPMRGYRQRYSLSEVGSSGLVSDANMAIARAGI 769
Db 523 SDIYHLSPT--LRTGYELKSKTSS--ITPAITPNDTDYILPRSGIT--ASTISLEYAGL 577
QY 770 SGVYSEGDNAVGSNRAHQMTGGIO--AGY-----IMSDNHNVPYRLRFPAG 815
Db 578 GGDQEF-----ISSSKFNFYQLODYGIDYIRYKASFYKVMDEGY--LPINQRIYLG 631
QY 816 DQSTIRGAHNSLSPISKGYLTGGQVLAAGTAENYEFMDLR--AVFGDIGANAYRKG 873
Db 632 IRSIRGESRTVSPKQMGGEIGGTIAFANSVELSPFLDIRIKIRGSGVFPDYGMGRKL 691
QY 874 TNDKIGAGVGMASPVGOVADVATGV--KEEGNPRIKLHFFIGTFE 919
Db 692 DEIKRMSTGICIEIPIPIGLQLVFAKPLMDKKDDINSEFNNGTRF 739

RESULT 20
B87448
conserved hypothetical protein CCI1603 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: B87448
R:Netman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: B87448
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-628 <STO>
A:Cross-references: GB:AE005673; NID:g13423000; PIDN:AAK23582.1; GSPDB:IGN0148
C:Genetics:
A:Gene: CCI1603

Query Match 4.6%; Score 216; DB 2; Length 628;
Best Local Similarity 21.8%; Pred. No. 0.00013;
Matches 110; Conservative 87; Mismatches 236; Indels 72; Gaps 17;

QY 456 DESTLEVETVELTGDILMDISPIEFASNLQDKLNL---VAAKARHLY-----D 504
Db 155 DEGVROAAVAMRLTEGEPSADYVGAEGRIYAQVAKLGVADVAAPREVYVADHARTV 214
QY 505 MPDDRVL-----INHDDGV-----NRSITGRISDAVSAVARALPDESENEVIDLEPRT 554
Db 215 RPFRIAGLVLRLNGVDVYTKGRTNPEWGRKLPWYA-----GDVYDPEQVA 262
QY 555 ALANRKTTPADYQSKV-----PLYVVASDKPRDQIGLGLGSGDTGRLV 600
```

```

Db 263 ELERRLDATAAYDISVSLAGTDKASAEGRYPVVVTLSDRRARTELEGAGSISEGACVD 322
QY 601 TKFEHNLINRDGYQAGAEIRLSEDKKGVLYATKPLSHPLNDQLRATLGXYQOEFHSTN 660
Db 323 ARWR--YNRCKRADITTYALRFALKLEORLGAELISLPHWRPQ--QTLKLKSSVFRNTD 378
QY 661 GFDISTPLEHEISRTIIONGKNRTYSLRYRLDKLTKQAPPETWODLPVFNKPSQE 720
Db 379 AYNETGATVGVDLTR--RQTTAYRTFVSPDLISQTEQVARN-----GLIGRKLN 429
QY 721 ALLAGVAVK--TVADNLVPMRGYRQRYSLSEVGSSGLVSDANMAIAR--AGISGYSEFGDN 778
Db 430 ATLLGLAIVAMDESDDIDLPKRGKRLETRAE--PTYAGDSVYPLKLAGGSAVL----- 483
QY 779 AYGSNRAHQMTGGIOAGIWSDNHNVPYRLRFPAGDQSTIRGAHNSLSPISKGYLTG 838
Db 484 PFGKQDSTVLAARVKLAILLGAGLLDVPAISRFPSSGGGSGVRYAYQAIGPRLSDNFPQ 543
QY 839 GOVLAAGTAENYEFMDRLAVFGDIG--NAYDKGFINDKIGAGVGMASPVGOVRY 896
Db 544 GISLVTSFEYRQKITDRWSGAVFVDAIGTHTPEQREDFRAGAGLVRYDLGFGPIRA 603
QY 897 DVATGV--KEEGNP--IKLHFFIGTFE 919
Db 604 DIAPLGRKKDPRKFOYLSIGQSF 628

RESULT 21
G81003
conserved hypothetical protein NMB2134 [imported] - Neisseria meningitidis (strain MC
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: G81003
R:etzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qiu, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masigian, V.; Piazza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappapoli, R.
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: G81003
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-635 <RPT>
A:Cross-references: GB:AE002561; GB:AE002098; NID:g7227384; PIDN:AAF42442.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB2134

Query Match 4.5%; Score 215; DB 2; Length 635;
Best Local Similarity 22.1%; Pred. No. 0.00016;
Matches 132; Conservative 73; Mismatches 225; Indels 166; Gaps 20;

QY 462 PVIEVETVELTGDILMDISPIEFASNLQDKLNLVAAKARHLVDMDDRVLAINHDDGVNR 521
Db 68 PDTESVKIKKPFVPLIQDQSEIKDMWEHPLITQOGEVLKQGTFLAEAPADNVKT 127
QY 522 SI--LGRISDAVSA-----VARAILPDESENEVIDLEPRTALAN 558
Db 128 MLRSKGYFSSKVSILTEKDGAYTHITPGRTKIANVGAVALIGDILSDNLAEYRNALEN 187
QY 559 RKTTPA-----DVYQSKK-----VPLYVVASDKP 582
Db 188 WQGVSGDFDQDSWENKTSVLGAVTKRAYFLAKLGNQAAVNPDTAFADLVVYVDSGRP 247
QY 583 RDQIGLGLGSGDTGTR-----LVTKFEHNLINRDGYQAGAEI 619
Db 248 ---IAGDFITGTQRPQEQIVSGIARFGPMPYDLDLDFQAL--EQNGHYSGASVQ 302
QY 620 ---RLSDKKGVLYATKPLSHPLNDQLRAT---LG-----YQGEVGHSTNG 661
Db 303 ADFRLQGDVRAVYVSVYEVKRRHLEGTIRLDSEYGLGKIAYDYNNLFNKGYIGSVYWD 362
```

```

0Y 662 FDLSTRLHEHISSTIIONGW---NRTYSLRALDKLTQAPPEPWODLPUVFWNGKPS 718
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 363 MDYETTLAAGISOPRNYRGNYWTSNVSYNRSTONLEKRAESGWY-----YVRDAG 416
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 719 QEALLA-----GVAV-----HRT-----VADILVNPFGY-----RQRY 747
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 417 IDARLGAELAEGRKPIGSAVDLGNSHATMLTASMKROLLNNVTLPHENGHYLDCKICTTL 476
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 748 SLEVGSSGLVSDANMALARAGISGYVSGFDNAYGGSNRAHQMTGT----OAGYIWSDNFN 803
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 477 GFLLSTALRTS---ARAG-----YFETPENKKLGTEFLIRQAGYTVARDNA 521
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 804 HVPYRLRFEGADOSIRINGADHSLSPISDKGYLTGGQVAVTAEYNEPMKDLRLAVFG 863
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 522 DVPSGLMFRSGGASVNGYEIDSLIGLAPNGSVLPERALLVGSLEYQLPPTRLTSAGVFH 581
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 864 DIGNAYDKGTNTDTKIAGVGYRNASPYGOVRVDYATGVYKEGCPNPKLHFETGTP 919
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 582 DMGDAAANFKRMKLKHSGLGVRFVSPFLAPESFIDAYGSDK--KIRWHISLCTRF 635
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

## RESULT 22

outer membrane antigen XF1046 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C:Accession: E82731  
C:Organism: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; PMID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A55328 below  
A:Accession: E82731  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-784 <STM>  
A:Cross-references: GB:AE003941; GB:AE003849; NID:g9105978; PIDN:AAF83856.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpton, A.J.G.; Relnach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, E.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir  
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.U.S.  
Submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferrio, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitaajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laiz  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Sanceli, R.V.; Sava  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A55328  
C:Contents: annotation  
C:Genetics:  
A:Gene: XF1046  
Superfamily: protective surface antigen D-15

Query Match	4.58;	Score 214;	DB 2;	Length 784;
Best Local Similarity	20.58;	Pred. No. 0.00025;		
Matches 177;	Conservative 120;	Mismatches 300;	Indels 266;	Gaps 45

[illegible]

QY	376	INTTDDKILPVKRELLLEQLLTVMGKAYNLQANRALSNLQIARRYNMNTELVFERRO-	434
Db	157	MTT---TTTPPLDNRNVDVTAIKEGKAARIH----NLIGKEKRNKDVSAMESSKHN	209
QY	435	-----IONDVS-----FEOSSSRTEPAVD-----ESTLEP-----VLEWELNDG	472
Db	210	WASWTRRDDQYSKEKLGSDLEKLNWYLDRCVVDNDISTOVISIEPEKHNMTTACVTEG	269
QY	473	ILMDISPIEFSANLI-ODKL-NLVAARAHLYDMPDDRVLAINHDDGVNRSILGRISA	530
Db	270	DQYKISSIKVTGNVLPOEKIEKLVIPKTDGIF-----SRVL-LEVSAAIINTLSNIGYA	324
QY	531	VSAAVARAILPDESENEVIDLPERLTALANKRPADYQSKYPL--YVVASDKPPDGOIG	588
Db	325	FSKVN---PIPTAKRA-----DRYAVN---LHYIPGRVYVROILFKGNRTSD---	368
QY	589	LGWSDGTGRVLTKEHN-----LINRGYQAGMELRL-----SDDKGVLY	631
Db	369	-----EVLRRROKRENSWYSOAIDRSKIRLQRLGYFEADVESTPVGSDVOYDI--VY	422
QY	632	ATKPLSHPLNDLRLATLGYOQEVFGHST-----NGFDLSTPTEHEISRTIQ-----	679
Db	423	TYKETT---SGSFQYGLGY- SKTYGVTTSVOLSQNNFLGSGNRVSVDSRSRYODRYSFS	478
QY	680	-----NG---GNRTY-----SLRYRLDKLTQ--APPETMODLPVDFVNGK	716
Db	479	YTNPFPTDNGVSLGYNLAYOKLDYSDFNAAQYNSRMSGOTIFGIPTEEND--TVSMVIGA	537
QY	717	PSQELLLAGVAVHKYVAD-----NLVNMGRGYROARYSLEY	751
Db	538	DSNOJTTTTPPGSTPKAIDYIDAVGQRTFRAMTELGMARDTENDYFMNLGMYORIGAEV	597
QY	752	---GSS-----GLVSDANNAIRAGISGYVSSGDNAYSNRHMOATG	790
Db	598	TLPGSTITYYKKNYISKYWPIIPALVINTRLTY-----GYGDD--YKGSHTRIILPD	647
QY	791	GIQAGYIMSDNENHPYRLREFAGDOSIRGAYHDSLSPISD-----KGYLTGQGVAV	844
Db	648	GTVA-----TASGLPFEENFAGCTNSVRGRDMDLPRBSRYVYALYNQOGPILGSPFTV	701
QY	845	GTAENYERMKD---LRLAVFGDIGNAVY--KGF-TNDFKIGAGYGVWASDPVGOVRVY	898
Db	702	GSTEMYFPKLFDSPESARISAFLEDFGNVFNQVNMFKANELRASSGVALLRAPIGISISY	761
QY	899	ATGVKEEGNP---IKLHFFIGTPEF	919
Db	762	AFPIKKENNDIETIQLTFEGQF	784

## RESULT 23

Probable surface antigen YPO1052 [imported] - Yersinia pestis (strain CO92)  
 C:Species: Yersinia pestis  
 C:Date: 02-Nov-2001 #sequence,revision 02-Nov-2001 #text,change 09-Nov-2001  
 C:Accession: AC0129  
 R:Parkhill, J.; Wen, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.  
 R:deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan,  
 I.L., M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:1158360  
 A:Accession: AC0129  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1795 <KUR>  
 C:Cross-References: GB:AL590842; PID:CAC89894.1; PID:g15979119; GSPDB:GN00175  
 C:Genetics:  
 A:Gene: YPO1052  
 ;Superfamily: protective surface antigen D-15

Query Match	4.58	Score 211.5	DB 2	Length 795
Best Local Similarity	19.18	Pred. No. 0.00036		
Matches 172	Conservative 139	Mismatches 306	Indels 283	Gaps 42

```
QY 148 NSEVVVPTLEPEKPG-LIKRLVA-----RLFNDGVNKKVPLKAKFYQSSQSGETSAT 199
DB 45 NMPVRVDDTSDDDIGKTIRALFATGNEFEDVRIDGNTLIVQK-----ERPTI 94
QY 200 GSSHQKTEPPIANAALEDITQESAMOLNGSIRPLRQALVAARAVGYDDLSITRR-- 257
DB 95 ASI-----TSGNKAHVAKEDMLKQ---NLEASGVHVEA-----LDRTTISNIE 134
QY 258 -----SIGEVVIIHDLGEPV---YIDRAVEVRGEA-----DDKAFTT----- 294
DB 135 KGLEDFYVSVKYSASVAKAVTPLPRNRVDLKLFTBESVAKIQOINTVGHSTTDLI 194
QY 295 ----VADEVPL--LIGDVFHGKETKKNLIENASAEHGYDGRMLRSVDVILPDNTAD 348
DB 195 SRFOLREVEPMWNVVGD---RKYOKOKLADDLLETLSFSYLDRCYARPNIDSTGVSLPD 250
QY 349 VSLIYD-----GTQYRDEVVFFITDPKTNQLTTPDPKLPVKRELLQOLLVNMGEAYN 403
DB 251 KKGIVTINITEGPOFKLNSIVY-----SGNLGHQSEAEKLTKEPGELEFN 297
QY 404 LQAVRALSNDL--IATRYFNWVNTETVPEREQIONDOVSFEQSSSRTEPAOVDESTLE 461
DB 298 GSKVTRBEDDIKKMLGRGYAVPRVYQPE---IND-----DDKTVK 336
QY 462 PVITVELTGDILMDISPIEFSASNLIQDKLNLVAARHLVD--MPDDRVLAINHDDGV 519
DB 337 --LHINVDAGNRPYVRHRIREFGNDTSKD--SVLRREMRQEGAMLGNDQVEA--GKERL 389
QY 520 NRSILGRISDAVSAVARAILPDESENEVIDLPERTALNRKTPADVOYSKKVP-----LYV 575
DB 390 NR--LGTF-----ETVDY-----ETQAVPGAADLVLD 413
QY 576 FVASDKPRDG---OIGLGWSDGTGRVLTKEFHNLINRQYQAGAEPLSEDKGKVLXA 632
DB 414 VTYVYKERNTGSLNFGIGCYGTESG---VSFOVGVOQDNMLGTGNTVGINGTKNDYQTYA 469
QY 633 TKPLSHPLNDLRLATLGVQOEVFHSTNGFDSLRTLEHEISRLSIQNGW--NRTYSRLR 691
DB 470 EFTLMDP-----FTYDGVSLGGRIFYNDKRADNADLSGYTNSSYG-- 510
QY 692 RLDKLKTQAPETWQDLVPDVFNGKPSQE--ALLAGVA--VHKTVADNL-----V 737
DB 511 -----ADGTILGPIINENSLRKYGVYVINDLSMLPQVAMRMYLESV 552
QY 738 NPMRGYROR-----YSLENG-----SSGLVSDANNAIARAGISGVY---SF 775
DB 553 GERPGYDGRGFTYDDFTLANIGMTYNNLDRGFFPTSGVKSVMNTRKITVPGSDNEFYKVTF 612
QY 776 GDNVY---GSMRAHQMTGIGQGYIMSDNFNHPVRLRFFAGGQOSTRGVAHDSLSP--- 829
DB 613 DTSAYOPLNEDRSVNLGRGLGIGDGIKSEMPYENFYAGSGSTVAGFSSNNIGPRAA 672
QY 830 -----ISDKGYLTGGQVAVAGTAE-----YNEEFMKDLRLAVFGDIGNAVYDKGFT 874
DB 673 YYANGCATVTNSTDAVGMAVAASIELITPTPISEKYSVSRSIRPIDSGTVDWTWME 732
QY 875 NDTK-----IGAGYVMAWSPVGOVRYDVATGYKE--EGNPI-KLHFTIG 916
DB 733 NTAKTRAAGIPDYCKASNIRVSAVALOMSPPLSVFSAKPVKDYEGDKSEQFOFMIG 792

RESULT 24
A:Accession: A13488
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-623 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL53076.1; PID:917983939; GSPDB:GN00190
A:Experimental source: strain 16M
C:Gene: BMEI1895
A:Map position: 1
Query Match 4.4%; Score 209; DB 2; Length 623;
Best Local Similarity 21.0%; Pred. No. 0.00033;
Matches 146; Conservative 79; Mismatches 263; Indels 208; Gaps 28;
```

```
A:Accession: A13488
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-623 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL53076.1; PID:917983939; GSPDB:GN00190
A:Experimental source: strain 16M
C:Gene: BMEI1895
A:Map position: 1
Query Match 4.4%; Score 209; DB 2; Length 623;
Best Local Similarity 21.0%; Pred. No. 0.00033;
Matches 146; Conservative 79; Mismatches 263; Indels 208; Gaps 28;

QY 310 KGYETKKNLIENASA-----EHG-----YFQGRW-----LD 335
DB 42 GKEADLKSVIEGASGLVSDADKPAQSAGLAKAGADYRRILSALYGGRRGCTISIKVD 101
QY 336 -RSVDVILPD-----NTADVSLYDGTQYRFEVVEFTIDPKTNQLTTPDPKLPVKRE-- 388
DB 102 GREANDIPDTEIPNNAKVALTYDPGQFLSRTAISNIAPEPCNRR--DKVOTPEBAG 158
QY 389 -----LLEOLLTVNKGAYNLQAVRALSNDLAFRYFMWVNTETVPEREQIQ 436
DB 159 FAPGGEAKSGTILKAERLAEAMWREGYAKAKAVTEGDVYADHADNRVSADILDPGRKAY 218
QY 437 NDQVSFEQSSSRTEPAOVDEST--LEPVIEYVELTDGILMDISPIEFSASNLIQDKNLV 495
DB 219 YGPVSV--VGTARMDPQVAVMTGLKPGQE----- 246
QY 496 AAKARHLTMDPDRVLATINHDDGVNRSILGRISDAVSAVARAILPDESENEVIDLPERTA 555
DB 247 -----YD-PDIDENA-----KKRLGRME-----VFRAMTPEEADK----- 275
QY 556 LANRKTADVYQSKRPVLYVYVYASDKPRDQOTIGLGWSDGTGRVLTKEFHN--NLINRDGYO 614
DB 276 -----IEPDGSLPITLNVQERKPRFPGAGAEYSTIDGCGVNSYMMHRLMLGR--- 322
QY 615 AGAELRLSEDKKGV-----KLXA-----TKPLSHPLNDLRLATLGVQOEVEFHN 657
DB 323 -GERLRFPKAVSGIGSGDENSEFDPKNYTYLLGASPAKRPVYPRDVFATVTLDAKREV--- 378
QY 658 STNGFDLSTRLEHEISRIIONGMMNRTYSLRYRLDKLKTQAPETWQDLVPDVFNGKP 717
DB 379 ----LDATET-----SINAKTFTQIFS-----DELSGLYANASQGHNVDDVDFPK- 421
QY 718 SQEALLAGVAHVKTADVADNLVYPMRGYRORYSLEVGSSGLVSDANNAIARAGISGVYSF-- 775
DB 422 -RDFTTAG-----LEGNLL-----YDSRNKRPDSSGFLYLVGN-----IOPFEFHY 462
QY 776 GD-----NAYSNNRAHQMTGGIGQGYIMSDNFNHPVRLRFFAGGQOSTRGVA 823
DB 463 GNFAITFTAEGRTHYHGFQOTDRVVLVLAGRLKVGSIYGGSIADLPQSOLFLAGGGGSGVRIG 522
QY 824 HDSLSPISDKGYLTGGQVAVAGTAEYNEEFMKDLRLAVFGDIGNAVYDKG---FTMDRTIG 880
DB 523 YRNIGVSANGNIIIGRSLVEANGAEVTRITDLSICAVAFVMDAGYGEKSFPPFSDQMYG 582
QY 881 AGVGYRWASPVGOVRYDVATGYKEEGNPILKHFITG 916
DB 583 VGGGLRYLTSLGPIRLDVAVPLNRRSGDPNRYGFYVG 618

RESULT 25
A:Accession: A13488
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-623 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL53076.1; PID:917983939; GSPDB:GN00190
A:Experimental source: strain 16M
C:Gene: BMEI1895
A:Map position: 1
Query Match 4.4%; Score 209; DB 2; Length 623;
Best Local Similarity 21.0%; Pred. No. 0.00033;
Matches 146; Conservative 79; Mismatches 263; Indels 208; Gaps 28;
```

A: Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A: Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
 A: Reference number: A64520; MIDID:97394467; PMID:9252185  
 A: Accession: G64601  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1-916 <TOM>  
 A: Cross-References: GB:AE000511; TIGR:HP0655

Query Match 4.4%; Score 206.5; DB 2; Length 916;  
 Best Local Similarity 19.3%; Pred. No. 0.00087;  
 Matches 202; Conservative 143; Mismatches 318; Indels 383; Gaps 51;

```

QY 24 MTSQALAOQNNPANIINHPAHDTAINQAKGNPVLPPEQIOAR-INAAGLNAPQSG 82
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 31 LTPSPKASQESQK---NEAPKNEVORNEAQKEFPQSQMOTKEKKVKSISVGLSYKMDL 87
QY 83 ALDVVAFDDQSPISRIQGEOSPGLDMSVIEETTPUSLEELFAOESTEMGINPDYIPEY 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 88 ANEIVKI-----RVGDIYDSKKIDTAVL-----ALFNQ----- 115
QY 143 QGEQNPSEVVPPTLEPEKGLIKRLYARLFNDGVNKKVPRLKAKFYOSSQSG--ETSATG 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 116 -----GTFKDYVA--TEEGGI-----LEHFPEKARIAGVEIKGYG 149
QY 201 SSHOKTEPYANIKALEDITQESAMDLSIPRLROTAL-VAARAVGYDIDLSIRNSI 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 150 TEKEKGLKSQMGIRKKGFDEBQKLE-----HAKTALKTALEGGGY----- 131
QY 260 GEVDYIHLGEPVYIDYRAVEYREGADKATTVAD--EVPLLIGDFHHGKYETKK 316
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 192 GSV-----VEVREKVEGALLIYFDVNRGDSIYIKQSYEGSAKLKR 224
QY 317 NLLENASAEH-----GYEDGR-----WLDKSVQ--VILPD 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 235 RMIESISAKQDEPMGMWGLNDGKRLDQLEYDSMRIDYWRKGYLDAHSSPFLKTD 294
QY 345 -MTADVSLIYDT--GYQRFDEVPFTIDPKTNQLTTPDKLTKRELLLEOLLTVNMGEA 401
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 295 FSTHDAKLHKYKKEGIOYRISDL-----IENDPVPVLK--TEKALKVKKRV 342
QY 402 YHLQAVRALSNL---IATRYFNMTVEIYFPEREQIQNDQVSFEQSSSRTEPAOVDES 458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 343 FNIELHRDAQILKTELAKGYAFA---VVKPDLDK-----DEK 378
QY 459 T-LEEVIEVELTDLGILMD--ISPIEFSASNLI-----QDKNLVAAKRAHLXMDP 506
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 379 NGLVAVITRIEVDWYIINDVYIISGNQRTSDRIIRRELLIGPKDKYNLTKLR----- 430
QY 507 DDRVLAINHDDGVNRSILGRISDAVASAVARAILPDESENEVIDLPERTALANKRPADYV 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 431 -----NSENSLR--LGFFSKVIEEKRV-----NSISDL----- 459
QY 567 QSKKVPYLVFASDKPRQQT--GLGWSDTGTRLVTKEFHNLIINRDGYQAQAEILRSED 624
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 460 -----LVSEEGRGICQLQFLGGLGSGYGGLML--NGSVSERNILFGTQSMISLVAN 506
QY 625 --KKGVLKATKP-----LSHPLNQLRATLIGYQOEVEGHSINGDLSTRLL- 669
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 507 IATGGGRSTPCGPKAGRMFAGNLSTLNP-----RIF--DSWYSSTINLY 549
QY 670 -EHEISRSIIIONG-----MNRTY-SLRYRLDKLKTQAPPETQDLPVDFNG-- 715
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 550 ADYRISTYQIOGGGFGVNVGRMLGNRTHVSLGYNLNTKLLGFSPLNRYTSYVNEYV 609
QY 716 KPSQALLAGVAHVKTVA-----DNLVNP-----MRCYROR-----YSLRV 751
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 610 SPROCSTPASYIINRLSGKTPPLQPESSCPAITSPEIRGIMPDYHTPTSSFTLDV 669
QY 752 G-----SSGLVSDANNAIARAGISGYVSPD--NAYGSNRKHAQMTGG----- 791
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 670 SYDNTDDYIFPRNGVIFS-----STATMSGLPSSGTLNMSWNLGQNVNWKYVYGRFAAHV 724

```

```

QY 792 -----IQAGYIWSDNF--NHVPYRLRFFAGDPQISIRGYAHDLSIPSDGY 835
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 725 HLGKYLILDLIARKEKTGGYIFRNTMDYILPNTFTYMGVTTVYRGRNSGVTPKDEFL 784
QY 836 LTTGQVLAAGTAENYEFMK--DIRLAVFDIGI-----N 867
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 785 WLGGDGIFTASTELSYGLVLAAKMRMLAMFDFGLTEKTPTRGSFFYNAPVTANFRDYG 844
QY 868 AYDKGFTNDT-KTGAQGVKRMASPVG 892
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 845 VIGAGFERATWRASTGQIEMISPMG 870

```

## RESULT 26

D71726

outer membrane protein omp1 (omp1) RP160 - *Rickettsia prowazekii*C:Species: *Rickettsia prowazekii*

C:Date: 21-Nov-1998 #sequence\_rev1sion 21-Nov-1998 #text\_change 03-Nov-2000

C:Accession: D71726

R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark

Nature 396, 133-140, 1998

A:Title: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.

A:Reference number: A71630; MIDID:99039499; PMID:9823893

A:Accession: D71726

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-766 &lt;AND&gt;

A:Cross-References: GB:AJ235270; GB:AJ235269; NID:93860572; PIDN:CAA14627.1; PID:9386

A:Experimental source: strain Madrid E

C:Genetics:

A:Gene: omp1; RP160

C:Superfamily: protective surface antigen D-15

Query Match 4.3%; Score 205.5; DB 2; Length 768;  
 Best Local Similarity 20.1%; Pred. No. 0.00074;  
 Matches 171; Conservative 111; Mismatches 297; Indels 271; Gaps 40;

```

QY 165 IKRLVLR-----LFNG-----VNKPYRLKAKFYOSSQSGERSAI-----GSS 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 65 IKRLVLSIFRNINMYITNDNLIVNTEPRTSSVFSNGSKIKTNILAKEITYMSGS 124
QY 203 HOKTEPYANIKALEDITQESAMDLSIPRLROTALVARAVGYDIDLSIRNSIGEV 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 LSGAKIELDVKKLE--IYKRSGRFTKVPKIK-----SLNNRV--- 163
QY 263 DVIHDLGEPVYIDYRAVEYREG--ADKATTVADVP-----LLIGDFHHGKYETK 315
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 164 -KVIFDIAEGPKTVIKSIYFSGNEHYSDELSKSTIVLTKESRMFRFLESNDYDPDRVEYD 222
QY 316 KNLLENASAEHGFEDRMLDRSDVILPNTADVSLIY--DTGTQYRFDEVPFTIDPKT 373
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 223 KELLREFYQSGFADFRVY--SASVALNDTKKEYFTTYSIEEEKYRFGNV--TID-- 274
QY 374 NQLTTPDKLPYRELLEOLLTVNMGAEVYLAQVRLASNDLITRFEN-----MVMTEIY 428
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 275 NKLITN-----INIKQLNKIVINKQKIKNMKTVDDIAEKI--GEFTANGPAVN---V 323
QY 429 FPEREQIQNDQVSFEQSSSRTEPAOVDESTLEPVLEFVELTDLGILMDISPIEFSASNL 488
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 324 YPD-----IKKNDNHADIKFIEIKADKYV 348
QY 489 QDKNLVAAKRAHLXMDPDRVLAITNHDGVNRSILGRISDAVASAVARAILPDESENEVY 548
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 349 INKINII--NNLKTEHDVIRAEKIEEGGVNMNRSTYRK-----GERMLRL 392
QY 549 DLPERTALNRKTPA--DVQSKKVPVYFVASDKPRDQIGLGWSDTGTRLVTKP---E 604
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 393 DYFEKYSTIATKAKDKYD-----VAVEYDEKSTSSIGDLGNITAGLFGRESFLE 445
QY 605 HNLINRDGYQAQAEILRSEDKGVKLYATKP--LSHPLNQLRATLIGY-----Q 651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 446 RNLVYG--TGKLLNAGVQVSKNSTSYGCLTDPHFLDRDLSLVNAPFNRYGRCASVLTNTD 504

```



```

OY 542 ESENEVYIDL-----ERTALANKRTPADVYQSKKVL-----YVFASDKPRCOI-- 567
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 365 VSENGVVTLLQVAGGVVENISVFRFRNKEGQDVNEQOQPIRGRTQDYITREVELKPKQOVFN 424
OY 588 -----GLGWSDTGTR 598
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 425 RNTVQKDLQVFEVTGLFEDYVNSLDPGDPTRKVVNVVNVVENS SGI AAGAGISSSGLE 484
OY 599 LVTKFEHNLINRDGYQGAELRLSEDKKVKRLYATKP-----LSHPLNDQLRATL-- 648
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 485 GWVSYQQOONLNGRNNQKLGAEVQLGERELFLDRFDPMIGGDPRTSYTANIFRRRSISL 544
OY 649 -----GYOOEV-----FCHSTNGFDLS--TR-----TLEHISRSTIIONGMWRTYLRR 692
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 545 ITDGKDEDITTEPDPGNPDNTWGQDPRTVRLGGVTEFRPRLSANPEREAWTASAGLOYO 604
OY 693 LDKLKTQAPP-----ETWQDLVPDFVANKPSQEALAGVAVHKTVADNLV 737
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 605 --RVSTRDADGNGLRKQDAVFDDNGNRRTSSEIPLSF--SGIGEDDLLVQLGQDRDLANNPL 661
OY 738 NPMRGYRQRYSLGYSGGLVSDANMAIARAGISGVY-----FGDNVGSNRAHOM 788
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 662 QPFGSGFLRRG--VDQSVPYSGNIFLTR--FRGSYQYLLPVKFTGF-----SKGPETI 711
OY 789 TGGIAGYIINDNPNHPRYRLRFEPAGDSIRGVNASHLSPI SDPGYLTGGOVLAVGTAE 848
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 712 ARIHQGYIVG-----LPPIEAFTLGGSNSVRYG-----EGCALSGNSFVQASVE 758
OY 849 YNTEFKMDLRLAVFGDIGNAVYDKGFTNDTK-----IGAGGVRRMASPVGOVREVDVA 899
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 759 YRFPVSVVSGALFEFDVGS--DLGTSRTAEVLNKS GSGYGLGVRVQSPIGPIRIDY- 815
OY 900 TGVKEEGNPIKLHPICTPF 919
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 816 -GINDGDS-RINFGIGERF 833

RESULT 29
F71910
  Probable outer membrane protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variatey: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
A:Accession: F71910
R:Alm, R.A.; Ling, L.S.L.; Mollr, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
  Ives, C.; Gibson, R.; Metberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
  Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
  A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: F71910
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-906 <ARN>
A:Cross-References: GB:AE001492; GB:AE001439; NID:g4155147; PIDN:AAD06181.1; PID:g415515
  A:Experimental source: strain J99
  A:Genetics:
  A:Gene: jhp0600

```

Query Match	Best Local Similarity	Score	DB 2	Length	906;
Matches	192;	Conservative	131;	Mismatches	311;
				Indels	333;
				Gaps	49;
OY	116	TPPLSEELFAOESN	-----EMGINPNDYIP-EYQ	-----GEOPNSEVVV	153
		: : :	: : :	: : :	
Db	37	TKREAKQNEKNETSOSNQTPKEKKVKSISTVGSYSMSDMLANETAKIRVGDMDVDSKKID			96
OY	154	PPTLEPEKDPGLIKRLVYARLFNDGVNKKVPRLAKAFYSSQSG--ETSALGSSHQKTPRYAN			211
		: : :	: : :	: : :	
Db	97	FAVLALNCGYFMDVYA-TEFNGI-----LEFHDEKARRIAGVELKGYTETREKEDDLKQ			150
OY	212	IKAALEDITQESAMDLDNGSLPRLKOTAL-VAARAVGYDIDLSITRNSIGEDVVIITHDLS			270
		: : :	: : :	: : :	
Db	151	MIKKKDDTDEQKLE-----NAKTLAKTALLEGQY-----GSV-----			184

QY	271	EYVYIDYRAVEYRGSGADDKATTYAAD---	EYPLILIGVYHHNGKETKKNLLENSASH-	326	
DB	185	-----VEETREKVESGALLIYFVDNRGDSIYIKOSIYEGSDKLRRVYTESLAKQ	235		
QY	327	-----GYFGR-----	WLRSDV--VILPD-NTADVSLTYD	354	
DB	236	RDFKMGMMGLNGKRLIDOLEYDSLRIDYVYRRGRIYDLAHISSPLFKLIDFTSHDAKLNHYK	295		
QY	355	T--GTQYRDEVEYFTIDPKTNQLTTPDPDKLPVKRELLBOLLTVNMGEAYNLOAVALS	412		
DB	296	VEKGIOYRISDIL-----	IEIDNPVVELK--TELEKALKYRKRVFNIEHRLADAQ	343	
QY	413	DL---IATFYEMVMTVEIYFPEREQIQNDQVSEBSSSRPEPAQVDEST--LEPIETVE	468		
DB	344	IKTELEADGVAFA---VYKPDLLK-----	DEKNGLVAKYIYIE	379	
QY	469	LNDGILMDSPIEFSASNLIQDKNLVAKARHLYDMPDR---	VLAINHDDGVRSILG	525	
DB	380	VGD--MVHINDYIISNGQRTSDRI-----	IRRELLGPRDKNMLKLNSESLAR--LG	430	
QY	526	RISDAVSAAVARAILPDESENEYIDLPERTALANKRTPADYQOSKVPPLYVYASKCPDPG	585		
DB	431	PEFSKYIEEKRY-----NSSLMDL-----	LVSVEEGRFG	459	
QY	586	QI--GIGMGSPTGTRFLYTKFEHNLIINRDQYOGAEILRSED--	KKGVKLYATKP-----	635	
DB	460	QLOFGLGIVSYGGLML---NCSYSEBNLEGTGSMISLYANLATGGGMSYRPMGPGAGRM	515		
QY	636	-----LSHPDLQDRLATLGLYQOEYFGHSTNGFDLSTRLL--	EHEISRIIIONG----	682	
DB	516	PAGNLSTLNP-----	RIF---DSWYSSTINLYADYRILSYOIQOQGGGFCVN	558	
QY	683	-----WNRFY--SLKRLDKLQTQAPPEWQDLPVUFVN--	GKPSOBALLAGYAVHK-----	730	
DB	559	VGRMLGNRTHVSLGYNLWTKLGLFSSPLNRYSSVNEVASPROCSPASYITNRLSGG	618		
QY	731	---TYADNLVNP-----	MRYGROR-----YSLEVG-----	SSGLVSDA	760
DB	619	KRPLVPESSSSSGCAITTSPEIKGIMDRDQYHPTYSSTFLDVSYDNTDQYPPRNQVITS-	677		
QY	761	NNAIIARAGISGYVSFGD---NAYGSNRHOMTGG---	10AG	795	
DB	678	-----SVATMSGSPSSGGLTNSMNGLGIVNVRNTKVYKFAAYHHLQKYLILDLIARPKTGGG	733		
QY	796	YIMSDNF--NHVYRKLRFPAQSDQSIGRAHDSLSYISDKGYITGGOVLAVGAELNYEFM	854		
DB	734	YIFRYNTDQYLRPLNSTFYMGVTVYVGRFNGSITPKDEGLMIGDGIJFTASTELSYVL	793		
QY	855	K--DLRLAVFGDIG-----	NAYDKGFINDT--KIDAGVY	885	
DB	794	KAAKKRIAMFFPFGGLTFRTPRGSEFYNAPTTANFMDYGVYAGGFERATIRASTGLQI	853		
QY	886	RNASPVG	892		
DB	854	EWISPMG	860		

RESULT 30  
AE2417  
hypothetical protein alr4893 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AE2417  
R:Kaneoko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, J.; Sasamoto, S.; Watanabe, A.; Iriyuzaki,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabatake  
DNA Res. 8, 205-213, 2001  
A:title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AE2417  
A:Status: preliminary  
A:Molecule type: DNA













```

Db      349 LRFEGTIVASDSTLRQEMROOECTWYNSQLVEGKIRLDRGFEETVENKIDPIINGSNDE 408
      565 ---VYOSKRVPLVYVAAADKPRDGOI--GLMGSDTGTRLVTFK-EHNLI----- 608
      409 VDVVYVVK-----ERNTGSIINFIGIGYGTESGISQASVKODNPLCTGAANSIAG 457
      609 NRDCYQAGAEELRLSE---DKKGVKLYATKPLSHPLNDQLRATIGYQOEVGHS-TNGFDL 664
      458 TKNDYGTSVNLGTYEBYFTFKDGVSLGNVFEFENYDNKSDTSSNKRTTYGSAVNTLGFVP 517
      665 SFRF-----LEH---EISRIITONGCMNRT---YSLRYRLDKLKTQAPPETWODLPVDFV 713
      518 NENNSTYYVGIQHTYKNKISNFALF---YNRMLYIYSKMFKNKNGIKTN-----DEDFS 565
      714 NGKPSOELLAGVAHVKTVDNLVNPBRGYRORYSLEPVSSGVLV---SDANMAIARAGI 769
      566 FGM-----NYSNLNRGIFEPFKGVKASLGGRYTIIGSDNKKYKLSADY 607
      770 SGVYSFGDNAVYGSNRH--OMTGGIOAGYIWSDNFHNVPYRLRFPAAGDOSIRGYAHDSL 827
      608 QGFYPL-----DRDHLWVYSAKASAGYANGFGNKRRLPFYQYTYTGGIGSLGFAFGSI 660
      828 SP-----ISDKGLTGGQVLAAGTAE--VYEFMKD-----LRLAVF 862
      661 GPNATYAEYNGSGTGTFKKISSD--VIGGMAITASAEILVPTPEVSDKSQNTVTSLSF 718
      863 GDIGNAY-----DKGFTNDTKIGAGVGYRMASPGVQVADVATGV 902
      719 VDAASVWNTKMKSDKNGLESVDLKRLPDYGKSSRIRASTVGVGOWOSPISPLPVSTAKPI 778
      903 K--EBSGNPIKLHFFIGTTPF 919
      779 KKYENDVDEQFOFISIGSF 797

Db
      39
      D70304
      Hypothetical protein aq_050 - Aquifex aeolicus
      C:Species: Aquifex aeolicus
      C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
      C:Accession: D70304
      R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oy
      V.
      Nature 392, 353-358, 1998
      A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
      A:Reference number: A70300; MUID:98196666; PMID:9537320
      A:Accession: D70304
      A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
      A:Molecule type: DNA
      A:Residues: 1-853 <AOF>
      A:Cross-references: GB:AE000671; NID:g2982793; PIDN:AAC06430.1; PID:g2982806; GB:AE00065
      A:Experimental source: strain VF5
      C:Genetics:
      A:Gene: aq_050

Query Match      3.6%; Score 171; DB 2; Length 853;
Best Local Similarity 19.4%; Pred. No. 0.082;
Matches 179; Conservative 136; Mismatches 325; Indels 284; Gaps 47;

      Oy      169 YARLFNDGVNKPRLKAKFYOSGSGEISAIGSHQKTEPYANIK-----AALEDITQ 221
      39 YAKIL-EMLKQPEIRREVYERKRDKLIYV-----BRYPIKIKIKGNVFLNDEDEV 90
      Db      222 ESAMDLSIPRLRGTALVAARAVGYDDLSIIRNSIGEVYIITHDGEV-YID---- 276
      91 KNVIGIEGIPLEIDNPRTYERILEKRYKEIGFL-NADRKVNISLDDKNAIFYINIRG 149
      Oy      277 --YRAVEYRGEGADDAKFTTVADEVPLILGDFVHHGKYETTKNLNLNENASAGYFGPGRWL 334
      150 NLVLEFDIQFESVKKLQKKELKASGLVIGSIFDIDKVDADAEENLENFRRKAGFFE---- 205
      Db      335 DRSDVILIPDNTADVSLIYDTG-----TQYRPDEVVFTIDPKTN-----QLTPT 378
```

```

      Db      206 -----SEVYILGKIKKELKRRFRALF-----PETDSFLKSLSIGLKNLVN 246
      Oy      379 DPDLKPVKRELLQ-----LLTYNMGNAVLYQVRAVNSLDLITRY-----FNMVTE 426
      247 HP--LATLKLALIGKVGWIPVYEVYEGERYEL--KFLGNFFSDELYLSLEFDINTVGVD 301
      Oy      427 IVFPE--REQIOND-----QVSEF-----OSSSRTPEAYV-----D 456
      302 ILVLEGKDKKEERYKRRKGFVDVAVVELEEDHRLILIKKEBERKAKAYINGKKEIETPYD 361
      Oy      457 ESTLEPVIE-----TWELTDGILMDISPIEESA-----SN 486
      362 EEKIODLSIKREIQYLEKMGVYATAYELKKINKDKKEVNVHVKINKGRYIWSFKIESE 421
      Oy      487 LIOD--KLVNVAAKAHILMDPDRVLAIN-----HDDGVNRSI-LGRIDANVASV 534
      422 LEFDLNRKISLKFPAKD-YETLDEIYKEIKRLREKGYFPAKVFPTDI RMQKVDALIMF 480
      Oy      535 ARA-ILPDE-----SENEVI-----DLPEPTALANRKTPEADYQ 567
      481 YKAVVITGEERYEGDTLITGLEKTRLKEAEVLEKEKTFSEVERSW--NAIESEITFK 538
      539 SLRLIEDYVDRSKKVRHLAYFOEKKRGVIGLSAGFNT-----FE-----GFKLSAE 584
      Oy      568 SKKVPVLYVFAADK-----PRDQIIGMGSDTGTRLVTFEHNLI NRDCYQAGAE 618
      539 SLRLIEDYVDRSKKVRHLAYFOEKKRGVIGLSAGFNT-----FE-----GFKLSAE 584
      Oy      619 LRLESDKKYKLYATKPLSHPLNDQ-----RATLGQOEVYFGSTNGFDLSTTLEHET 673
      585 LSL-RNLIGLIGLINTFNFSISEKYLEVSPKDNFLPSRRFPGETSLFKDYE---EHD 639
      Oy      674 SRSIIIONGWNRRYSRLRYRLDKLKTQAPPETWODLPVDFVNGKPSQ--EALLAGVAHVKT 732
      640 YELFTE---GFAFSGLRYK-----PTTFGIGISNFEAKTTGERDKGT 681
      Oy      733 ADNL-----VNPBGYRORYSLEVSSGLVSDANMAIARAGISGYVSGDNAAG 781
      682 YRKLGISFSYKERPRFLVLPFGYGHRSKIIELEKIK--KEFEKFGSRLKFSVG---YA 736
      Oy      782 SNAHQMTGSIQIACGYIWSDNFHNVPYRLRFPAAGDOSIRGYAHDSL-SPIISDKYLTGQ 840
      737 TKRA-----PIFERPFLGGYKMKGTYESISPL-----GGR 769
      Oy      841 VLAVGTAEVYEFPMKDLRLAVF---GDIGNAYDKGFNDPK-IGAGVGYRMASPGVQVRY 896
      770 QMLYISPIEIRYLLANRNLIELITFLELKVENKPSLYKMKKIDIGSAGFR--TPVGLIRG 827
      Db      897 DVATGVKE-EGNPIKLHFFIGTTPF 919
      828 DIAVPLEDFKVKPSRLKFLYLSVEF 851

      Db
      828 DIAVPLEDFKVKPSRLKFLYLSVEF 851

RESULT 40
AD2445
Hypothetical protein all5116 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AD2445
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2445
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-588 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA076815.1; PID:g17134254; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all5116
```



GenCore version 5.1.4-P5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:15:24 ; Search time 28 Seconds

(Without alignments)  
1361.312 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 4727  
Sequence: 1 MSKPVLFANRSEFPVALAAY.....TGVEKGNPKLHFFIGTPEF 919

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	436.5	9.2	578	YTFM_HAEIN	P44038 haemophilus
2	433.5	9.2	577	YTFM_BCOLI	P39320 escherichia
3	180.5	3.8	793	D153_HAEIN	O32629 haemophilus
4	178	3.8	810	UP05_BCOLI	P39170 escherichia
5	171.5	3.6	797	D151_HAEIN	P46024 haemophilus
6	167.5	3.5	795	D152_HAEIN	P44935 haemophilus
7	163	3.4	1325	BCC3_ACEXY	O9w763 acetobacter
8	159.5	3.4	1319	BCC1_ACEXY	P19450 acetobacter
9	159	3.4	475	YDF6_SCHPO	O10478 schizosacch
10	150.5	3.2	991	SCA4_RICAP	O9a164 rickettsia
11	149	3.2	1022	ACF7_HUMAN	O9u249 rickettsia
12	138.5	2.9	5430	ACF7_HUMAN	O9u249 rickettsia
13	137.5	2.9	495	G6PD_STRPN	O54337 streptococc
14	137	2.9	900	IF38_ARATH	O49160 arabidopsi
15	136	2.9	900	IF38_ARATH	O49160 arabidopsi
16	134.5	2.8	6486	TYCC_BACBR	O30409 b tyrocidin
17	134	2.8	1279	BCSC_PSEFL	P58937 pseudomonas
18	133.5	2.8	1036	P200_MYCPN	P75211 mycoplasma
19	133.5	2.8	1296	ASAL_ENTFA	P17953 enterococcu
20	133.5	2.8	4349	FAT2_HUMAN	O9u9q8 homo sapien
21	132.5	2.8	3726	TRX_DROME	P20659 drosophila
22	132.5	2.8	856	CLPB_HAEIN	P44403 haemophilus
23	132.5	2.8	1957	YD86_SCHPO	O10411 schizosacch
24	131.5	2.8	612	HS75_KLUMA	O9h178 kluyveromyc
25	131.5	2.8	896	RA50_THEAC	O9h178 kluyveromyc
26	130	2.8	1434	PROB_UREPA	O9p9v6 ureaplasma
27	130	2.8	981	SCA4_RICPE	O9a137 rickettsia
28	129.5	2.7	2738	PCGV_RAT	O9a137 rickettsia
29	129	2.7	1616	SLAP_BACCI	O38824 bacillus no
30	129	2.7	1115	FRPA_NEIME	P55126 neisseria m
31	128.5	2.7	1185	MAPX_DROME	P23226 drosophila
32	128.5	2.7	880	DP01_BACSU	O34996 bacillus su
33	128.5	2.7	999	DSG3_HUMAN	P32926 homo sapien
			2869	RBP1_PLAAB	O00798 plasmidium

34	126.5	2.7	1302	FRPA_NEIME	O9K09 neisseria m
35	126.5	2.7	1326	BCC2_ACEXY	O82861 acetobacter
36	126.5	2.7	1829	FRPC_NEIME	O91y5v neisseria m
37	126.5	2.7	2136	RPI_HUMAN	P56715 neisseria m
38	126.5	2.7	2198	YLAJ2_CABEL	P34367 caenorhabdi
39	126.5	2.7	3396	PGCV_HUMAN	P13611 homo sapien
40	125.5	2.7	1616	P200_MYCCE	O49429 mycoplasma
41	125	2.6	928	HXA2_HAEIN	P45354 haemophilus
42	125	2.6	963	TRES_THETH	O06458 thermus the
43	125	2.6	1433	REST_CHICK	O42184 gallus gall
44	124.5	2.6	3433	UTRO_HUMAN	P46939 homo sapien
45	124.5	2.6	3712	LMA_DROME	O00174 drosophila

## ALIGNMENTS

RESULT 1  
YTFM\_HAEIN  
ID YTFM\_HAEIN STANDARD; PRT; 578 AA.  
AC P44038;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Protein HI0698 precursor.  
GN HI0698.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RP [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
rd.";  
RL Science 269:496-512(1995).  
RN [2]  
RP IDENTIFICATION BY MASS SPECTROMETRY.  
RX MEDLINE=20137488; PubMed=10675023;  
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,  
RT Gray C., Fountoulakis M.;  
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";  
RL Electrophoresis 21:411-429(2000).  
CC -1- SIMILARITY: STRONG, TO E.COLI YTFM.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: 032752; AAC2357.1; -  
DR TIGR: H10698;  
DR Interpro: IPR000184; Bac\_surfAg\_D15.  
DR Pfam: PF01103; Bac\_surface\_Ag; I.  
KW Signal; Complete proteome.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 578 PROTEIN HI0698.  
SQ SEQUENCE 578 AA; 65812 MW; 34P9AC189C505876 CRC64;  
Query Match 9.2%; Score 436.5; DB 1; Length 578;









```

DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_TaxID=727;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype B;
RX MEDLINE=95255676; PubMed=7737523;
RA Plack F.S., Loosmore S., Chong P., Thomas W.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of
RL Haemophilus influenzae."
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Minn, and Eagan / Serotype B;
RX MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease."
RL Infect. Immun. 65:3701-3707(1997).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U13961; AAA85645.1; -
DR EMBL: U60832; AAB61974.1; -
DR EMBL: U60833; AAB61976.1; -
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
DR Antigen: Outer membrane; Signal.
FT SIGNAL 1 19
FT CHAIN 20 797
FT POTENTIAL.
FT PROTECTIVE SURFACE ANTIGEN D15.
SO SEQUENCE 797 AA; 87675 MW; 2F93DE538696AF1B CRC64;

Query Match 3 6%; Score 171.5; DB 1; Length 797;
Best Local Similarity 19.8%; Pred. No. 0.046;
Matches 158; Conservative 121; Mismatches 305; Indels 215; Gaps 39;

QY 263 DVIHDLGEPYIDYRA-----VEVRGEKA-DKRAFTVADEVPLLDGVFHHGKYEIK 315
DB 72 DVKAHQEGVLVSVVAKSIIDVKTKGNSVPTKLNKLNLDNGKVGVDL-----IREK 127
QY 316 KNLIINASAEHGYPFGRLIDRSVDVI--LPNTADVSLIYPTGQYRFDEVVEFTIDPK 372
DB 128 LNEFAKSYKEHAYSGRY-NATVEPIVNTLPNNRAEILIQINDEDKAKLASLTF-----K 181
QY 373 TNLQITDDPKLPVKRELLEQLTLTVKGEVY-----NIQAVRAISNDLIATRYEMVNT 425
DB 182 GNE-SVSSSTLOEQWELPDPSWMLKMGKFEGAOFEKDLQSTIR-----DYLLNGYA 232
QY 426 EIVPEREOIQDOVSEFQSSSRTEPAQVD-----ESTLEPYIEVETLDTG 472
DB 223 KAQITKTVOGLNDEKTKYNTIDVNEGQYQYDLSARITGNLGMSELEPLLSALHLNDT 292
QY 473 I-LNDISIEFSASNLIDKLVLAARHLYDMD--DVLAINHDDGVNRSIL-- 524
DB 293 FRRSDIADVE-----NAIKAKIGERGYSATVSPDPDANKTLAITLVVDAGRILTVNQ 348
QY 525 -----GRISDAVASVARALPDES-----ENEVIDLP-----ALANRKTTPAD----- 564
DB 349 LRFEGNTVASDSTLKQEMKROEGTWYNSQLVELGKIRLDRTGFEFTVENRIDPINGSND 408

```

```

QY 565 ---VYQSKVPLVYFVVASDKRRDGOI--GLWGSPTGTRLYTKF-ENHLI----- 608
DB 409 VDVIYKVA-----ERNKTSINFGISYGTGESGISYQASVKODNPLTGAANVTAG 457
QY 609 NRDCYQAGAEPLRSE---DKGVKLYATKPLSHPLNDQLRAFLGYQGVFGHS--TNGD 664
DB 458 TKNDYGTSVNLGYTEPYPTTKGVSLGKGVNFEENTDNKSDTSSNKRTTYSNVTLGFPV 517
QY 665 STRT-----LEH-----EISRSIIQNGWNR--YSLRRLDKLTKQAPETWQDLPVGFV 713
DB 518 NENNSYVYGGLGHTYKNISNFAL--YRNLYIQSMKFKKGIGITN-----DPEFS 565
QY 714 NGRPSQELLAGVAVHKTVDNLVNPARGYQQRISLEKSSGLV---SDANMALIARAGI 769
DB 566 FGW-----NYSNLNNGYPTPTGKASLGRVYTPGSDKKYKLSADV 607
QY 770 SGVYSFGDNAYGSNRH--QWTGGIOAGYISDNPNHVPRLRFPAGDOSIRGAHDSL 827
DB 608 QGFYPL-----DRDLHWYSASAKASAGYANGFGKRLIPYQTYAGIGSLRGPVAGSI 660
QY 828 SP-----ISDKGYLTGGQVLAAGTAE--YNYEFMKD-----LRLAVF 862
DB 661 GPNALYAEYNGSGTGFFKRISSD--VIGGNAITFASALVLPPTPYSDKSQNTVRSLSF 718
QY 863 GDIGNAY-----DKGFTNDTKIGAGYGVWASPVQGVRVDAATGV 902
DB 719 VDAASVNTKRWKSDKNGLIESDVLKRLPDKKSRIRASTGVGFQMGSPIGFLVSAYAKPI 778
QY 903 K--EGNPKLHPFGPFP 919
DB 779 KRIENDVEQFQSTIGSF 797

```

```

RESULT 6
D152_HAEIN
ID D152_HAEIN STANDARD; PRT; 795 AA.
AC P44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
GN HI0917.
OS Haemophilus influenzae.
OC Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shetty R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weiman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uitterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

CC or send an email to license@isb-sib.ch).

DR EMBL: U33773; AAC22575.1; -

DR TIGR: H10917; -

DR InterPro: IPR000184; Bac\_surfAg\_D15.

DR Pfam: PF01103; Bac\_surface\_Ag\_1.

KW Antigen: Outer membrane; Signal; Complete proteome.

FT SIGNAL 19

FT CHAIN 20 795 POTENTIAL.

FT SEQUENCE 795 AA; 87478 MW; B85691FC22E6ED44 CRC64;

Query Match 3.54; Score 167.5; DB 1; Length 795;

Best Local Similarity 19.68; Pred. No. 0.076; Indels 209; Gaps 39;

Matches 156; Conservative 123; Mismatches 307;

CC 263 DVIHDLGEPYIDYRA-----VEVRGEG-ADDKAFTTVDEVPDLIGDVFHNGKYEK 315

DB 72 DYKHHQGDVLYSVVAKSIISDVKKIKNSITPEALKQNDANGFKYGDV---IREK 127

CC 316 KNLINASAHEGYDGRWLDRSDVY---LPDNTADVSLYDTGTQYFDEVEFTIDPK 372

DB 128 LNEFAKSVKHEHVASGRY-NATVEPIVNTLPNNRAELIQINEDDKAKLASLTF-----K 181

CC 373 TNOITTPDPKLPVARKELLEQLLTVNMGEAY-----NLQAVRALSNDLIATRYPNMVT 425

DB 182 GNE-SVSSSTIQEOMELQPDSSMKLMGKFECAQFEKDLQSIK-----DYIANGYA 232

CC 426 EIVPEREIQINDQVFEQSSSSRTEPAQVD-----ESTLEPIETVELTDG 472

DB 233 KAQITKTDVQLDERTKVNVTIDVNEGLODYDLRSARITGNLGMKAELPELISALHLNDT 292

CC 473 I-LMDISPIESANLQDKNLVAAKARHLXMDPD---RVLAINDDGVNREIL--- 524

DB 293 FRRSIDLADVE---NAIKAKLGERGYSATVNSVPDEPDANKTLATLVDAGRSLTVRQ 348

CC 525 ---GRISDAVASAARAILPDDES---ENEVIDLP---ERT---ALANRKTTPAD--- 564

DB 349 LRFEGNYSASOSTLRQEMRQEGTMYNSQLVELGKIRLDRTGFETVENRIDPIGSMDE 408

CC 565 ---YQSKKVLVLYVVASDKPRDQI--GLMGSDTGTRLVTKF-EHNLI----- 608

DB 409 VDVVYKVK-----ERNTGSINFGIGYTESGISYQASVAKQDNEFLGTGAASVIA 457

CC 609 NRQGYQAGAEILRSE---DKKGVKLKATKPLSHPLNDQLRATLIGYQOEFGHS-TNGFDL 664

DB 458 TKNDYGTSVNLGYEPFTTKDGVSLGCVNFPFENDNSKSDTSSNKRTTSGSNVTLGFEPV 517

CC 665 STRT-----LEH---EISRSIIIONGMNRT---YSLRYRLDKLAKTQAPPETMODLPVDFV 713

DB 518 NENNSYVGLGHTYNNKISNFALE---YNRNLXIOMKFKGNGIKTN-----DPDFS 565

CC 714 NGKRSQELLAGVAVHKTVADNLVNPMPGYSRQRYSLLEVSGSLV---SDANMALARAGI 769

DB 566 FGW-----NYSNLNGYFPPTKGVKASLQGRVITPGSDMKYKYLSDAV 607

CC 770 SGVYSFGDNAGYSNRH--OMTGGIOAGYIWSDNFNHVPYRLREFPAGDOSIRGYAHSLSL 827

DB 608 OGCFPL-----DRDHLVYVSAAKAGYANGFGNKRLEPFOTYTAGISGRAGVYSI 660

CC 828 SP---ISDKCY-----LTGGQVLAVGTAE--YNYEFMKD---LRLAVFGDIG 866

DB 661 GPNAVYAEHNGCTFKKISSDVIGNAITVASAELIPLPTEFVSQKSONVTRTSLEFVAA 720

CC 867 NAY-----DKGFTNDKIGAGVVRBASPGQVRYVAVATGVK---E 904

DB 721 SVMTTKKSKSGSLDNNVLAFLPDYKSSRIASTGTVGQWSPIGLAVSTAFKIKYE 780

CC 905 EGNPIKLHFFIGTPT 919

DB 781 NDDVEQFOFSTIGSGF 795

RESULT 7

BCC3\_ACCEXY ID BCC3\_ACCEXY STANDARD. PRT: 1325 AA.

AC Q9WXB63;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Cellulose synthase 1 operon protein C precursor.

GN BCSCI.

OS Acetobacter xylinus.

OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;

CC Gluconacetobacter.

OX NCBI\_TaxID=28448;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JCM 7664 / IFO 13693;

RX PubMed=10382968;

RA Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M., Inoue Y.;

RA Inoue Y.;

RT "Cloning of cellulose synthase genes from Acetobacter xylinum JCM 7664: Implication of a novel set of cellulose synthase genes.";

RL DNA Res. 6:109-115(1999).

CC -I- FUNCTION: Required for maximal bacterial cellulose synthesis. It may be involved in the formation of a membrane complex for extrusion of the cellulose product (By similarity).

CC -I- EXTENSION: Bacterial cellulose biosynthesis.

CC -I- SUBCELLULAR LOCATION: Outer membrane (Potential).

CC -I- SIMILARITY: BELONGS TO THE ACSC/BCSC FAMILY.

CC -I- SIMILARITY: CONTAINS 8 TPR REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by, and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AB015802; BAA77587.1; -

DR InterPro: IPR003921; CellSynth\_C.

DR InterPro: IPR001440; TPR.

DR Pfam: PF00515; TPR; 4.

DR PRINTS: PR01441; CELLSTHASEC.

KW Cellulose biosynthesis; Repeat; TPR repeat; Membrane; Outer membrane; Signal.

KW SIGNAL 1 30 POTENTIAL.

FT CHAIN 31 1325 CELLULOSE SYNTHASE 1 OPERON PROTEIN C.

FT REPEAT 50 83 TPR 1.

FT REPEAT 85 117 TPR 2.

FT REPEAT 292 325 TPR 3.

FT REPEAT 406 439 TPR 4.

FT REPEAT 508 591 TPR 5.

FT REPEAT 702 735 TPR 6.

FT REPEAT 737 769 TPR 7.

FT REPEAT 737 769 TPR 8.

SO SEQUENCE 1325 AA; 142541 MW; DA4EECEGFADED74D CRC64;

Query Match 3.44; Score 163; DB 1; Length 1325;

Best Local Similarity 21.88; Pred. No. 0.29; Indels 284; Gaps 56;

Matches 221; Conservative 129; Mismatches 381;

CC 43 PAHDPAIN--QAKAGNPVLLTPEDIOARLNAGI-NAPQ-----SQAL----- 84

DB 165 PPHSLAIEVYQTMGVP-----ADWDQARAGLAGLVANPQYRAQLARAFQILTYMTSR 219

CC 85 -----DVVNFDDQSP--SRIGQSPPLGLD-HSVIEETPLSLIEELFAQESTEMGIN 134

DB 220 MEGLARLKDLOGFRTQAPVEAAAAAOSYROTJSMLEPVAETQPLMQOWLTA-----H 271

CC 135 PNDYIPEXQGEQSENVSEVVPPLPEPEKPLIKRLVARELNDQVNVKVPRLKA--KEYQSS- 191

DB 272 PDD-----TALKREHMLHPGGPPDKAGLARQAGFQQLNSG-----RLSADQSQSL 319

CC 192 --QSGETSAIG-----SSHQ-----KTEPYANIKAALEDITQE-----SAMDLNGSIPRL 234

```

Db 320 QINSHADSLGGLVSMRGDAEARRRYOEAMADPKTADRRPALAGMEISGDYAAV 379
QY 235 RGTALVA-----ARAVGY---DIDLSTIRNSICEVDYIHDLEPEYIDY 277
Db 380 RQ--LTAHQYTEAKORLSLARQPGFPGATMLADLOTTQOIDASEQ-----EX 429
QY 278 RAVEVR-----GEGADDKAFTTVADVEPLL--IGDFVHHKKYETKNLIE---NA 322
Db 430 RSLADPNNQALAMGLIARDMAOGNTAEARQLSRVGP-----QYATEVEGETVGLMA 484
QY 323 SAHEGFDGRMLDRSDVILPONTADVSLIYDGTQYRFEDEVFFITDPTNOLITDPDK 382
Db 485 AASH-----TSDSARKVALIREMTQAPRDPWVRIML---AALQOQGDV 526
QY 383 LPVKRELLBOLLTVNMGEAVNLQA---VRALSNDLIATRYFMVMTVEIYF-----ER 432
Db 527 AEAGRYM--QPLIANVYTQDRQAGILYTYGAGNDAATRLLSGLSEDEXSPAIRSIAE 584
QY 433 EGIQNDQVFEOSSSSRTEPAQVDESTLEPVETVELFDGIIMDISPIEFASNLIDOKL 492
Db 585 MOIKEDLAS---RLSWVPNPVPLIREALAP---PPPTGARGVAVADLFROGDMITHARM 637
QY 493 NLVAAKARILYDMPDDR-----LAINDDGYNRSILGRISD-AVSAAVARALLPDESE- 544
Db 638 ALRIASTRTIDSPDQRLAYATEYKMSNPVAAAR--LLAPLGGSGSGAGNALLPEQOQT 696
QY 545 ----NEVIDLPERTALANRKPADVYOSKKVPPLYVFAVSPDKPRDQIGLGWGSDDTGRIL- 599
Db 697 LQQLRMGIIVASDILLNQGDQAQAYDLHAPALRADPEATSPRLALARLYNGEGKSKAL 756
QY 600 ---VTKFEINLINRQGYQAGAEIRLSEDKKGVLYATKPLSHPLDOLRATLGY---OO 652
Db 757 DIDLAVLRHNPDDLRAQVAAVQAAVNSGRKSLATHLAMDQVQSPMDARWMLGMAVADQA 816
QY 653 EYFGHST---NGPDLSTRTLEHESRSTIQGWNRTYSLRYRDKLKTQAPPET---704
Db 817 DGHGHTTADLRAYDLRLQOQVEGSRAS---GPAATEE---DAL---APPSNPP 863
QY 705 ----MODLPVDFVNGKPSQEA-----LAGVAVH--KTVDNL---VNPMSGYRQ 745
Db 864 RHNGYGRQTELPARTGGSYSMEATSPLEAADOMLSSISOINTLRKRLAPSIDGGLGFRS 923
QY 746 RYSLVSGSGGLVSDANMAIA-----RAGISGYVSFGDNAYGNSRAHQMTGGIOAGYIWS 800
Db 924 R--SGEHG--MGRLTENIPIVGRPLQAGES-----SLFFSTPTMIWGS 965
QY 801 NFN-----HVPYRLRPFAGDQDSIRGYAH--DLSLPSIDKYL-----TGGVLANGTAE 848
Db 966 DLNAGSYDVP---RY--GTNATEAYINQYVNSLSQNNSSSSSLRTQOIOGGQ----- 1012
QY 849 YNFEFMKDLRLAVFGDIGNAYDKGFTNDKIGAGVGNWASPVGYGVAVNATGVK 903
Db 1013 -----GEAGFAPDYQFSN--SWVRADVG---ASPIGFPITTVLGVGE 1049

```

```

RA Wong H.C., Fear A.L., Calhoun R.D., Eichinger G.H., Mayer R.,
RA Anilkam D., Benzinman M., Gelland D.H., Meade J.H., Emerick A.W.,
RA Bruner R., Ben-Bassat A., Tal R.;
RT "Genetic organization of the cellulose synthase operon in Acetobacter
RT xylum";
RL Proc. Natl. Acad. Sci. U.S.A. 87:8130-8134(1990).
CC -1- FUNCTION: Required for maximal bacterial cellulose synthesis. It
CC may be involved in the formation of a membrane complex for
CC extrusion of the cellulose product.
CC -1- PATHWAY: Bacterial cellulose biosynthesis.
CC -1- SUBCELLULAR LOCATION: Outer membrane (potential).
CC -1- SIMILARITY: BELONGS TO THE ASCC/BSCC FAMILY.
CC -1- SIMILARITY: CONTAINS 9 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37202; AAA21886.1; -.
DR PIR: C43735; C43735.
DR InterPro: IPR003921; Cellsynth.C.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR: 3.
DR PRINTS: PR01441; CELLSENTHASEC.
KW Cellulose biosynthesis; Repeat; TPR repeat; Membrane; Outer membrane;
KW Signal.
FT SIGNAL 1 46
FT CHAIN 47 1319 POTENTIAL.
FT REPEAT 47 82 CELLULOSE SYNTHASE OPERON PROTEIN C.
FT REPEAT 84 116 TPR 1.
FT REPEAT 291 324 TPR 2.
FT REPEAT 325 338 TPR 3.
FT REPEAT 405 438 TPR 4.
FT REPEAT 557 590 TPR 5.
FT REPEAT 701 734 TPR 6.
FT REPEAT 736 768 TPR 7.
FT REPEAT 1059 1094 TPR 8.
FT REPEAT 1319 1414 TPR 9.
SQ SEQUENCE 1319 AA; 141431 MW; 48C6BEAFC6359A7F7 CRC64;

Query Match 3.4%; Score 159.5; DB 1; Length 1319;
Best local Similarity 19.2%; Pred. No. 0.45;
Matches 201; Conservative 131; Mismatches 360; Indels 357; Gaps 51;

QY 43 PAHDTAI-----NQAKGNPVLYLT-PEQIOARL---NAAGLNARQSQAL- 84
Db 164 PPHSLAVEYQYQWAGVPAGMDQARAGIAGVVASNPQDYRAQLAFQAQLTYNTSTRMGLT 223
QY 85 ----DVNPFDDQSPIT-SRIGSQSPPLGLD--MSVIEETPLSLBELFAQESTEMKINPDYI 139
Db 224 RLKDLQSFQARVEAAMAAQSYRQTLSQLPVNPEQOPLMEQMLSA-----HPND-- 273
QY 140 PEYQGEQPNSEVVVPPLEPEKPGILIKRLYARLENDGVNKPVRPKAKFYQSQSGESAT 199
Db 274 ----TALKREHMLHPGPGPPDKAGLARQ-----AGYQLAAGRIAAA 310
QY 200 GSSHQKTEPYANIKAALEDITQESAMDLNGSIPRLROTALVAARAVGYDIDLSIIRNSI 259
Db 311 EQSFQ-----SALQINS-----HDAD-----SL 328
QY 260 GEVDYIYIHLGEPVYIDYRAVEVRGEGADKATTVADVEPLLIGDFVHHKKYETKNL 319
Db 329 GKGGLVSMRGQDA--EARRYFEEMAADPK-----TADRRRPALAGNAVSGEYASVROL 382
QY 320 ENASAEHGYFDGRW-----LDRSVDT----- 341
Db 383 ----AAHQYTEAKOQALATLARQPGYTGATIMLADLQRSIGQIAAADBOEYRGILSREPNN 438
QY 342 -----LPDNTADV-SLIYDTGTQY--RFDEVVFFITDPTNOLITDPDKLAPVK 386

```

```

Db 439 QALMLGLARVDMAGNTAEAROLLRSVGPQYASVGEIEVSGLMAAASQTSDSARKVSTL 498
QY 387 RELLEQL-----LTVMNGEAYNIQ-----AVRAL-----SND 413
Db 499 REAMAQAPRPPWVRWVIRINLANALQOQGDVAEAGRWVQPIIANPVTAQDRQAGILTYTGSSND 558
QY 414 LIATRYNMYNTEIVPEREIOQNDQVSEFGSSSRTEPAQVDESTLEPIETVELDGI 473
Db 559 AMTROLLAGSPADYSPAIRSIA-EEMEIKODLASRLSMVSNPVLIREALTQDPPTGAR 617
QY 474 LMDISPIEFASNLIDQKLNIVAKARHLMDPDRV-----LAINDDEVNSIIGRI 527
Db 618 GVAADLPFRQGDVHARMALIRIASTRIDISLPQQRISYATEYMKISPVAAAR-LIAPL 676
QY 528 SDAY-SAVARAILPDESE-----NEVIDLPERTALANKRTPADYQSKVPPLYVFVASDK 581
Db 677 GDCGSGATGSAALLPEQYOTLQQLRMGISTVAGSDLLNRGDAQVADHLAPRLQADPEATS 736
QY 582 PR-----DGOIGLQMGSDTGTRLVTKFEHNLINRGYQAGAEIRLSEDKKGV--KL 630
Db 737 PKLALARLYNGHGKPKALEID-----LAVLRHNPQDLDAHOAAVQAANSNDHNSLATRL 791
QY 631 YATKPLSHPLMDQLRATLGYQGEVFGSTNGFDLSTRLEHISRSIIQNGMWNTYSLR 690
Db 792 AMDCVQSPMDARAMLMAVAADQDGH-----GQRTIE-----DLRRANDLR 833
QY 691 YRL-----DKLKTQAPPET-----MODLPDVFYNGKPSQEA----- 721
Db 834 LQVEGTGTRAAAGAGAOEDAL--APPSTNPFPRGYGHQTELAGAPYTGSGYSMAASPD 890
QY 722 -----LAGVAVH-KTVADNL--VNPMRGYRQRTSLEYSGGLVSDANMAIA-----RAG 768
Db 891 TSDMLSTISGQITTLRENLAPSIDGLGFRSR-SGEGH-MGRLEAPIPIVGRLLPQAG 948
QY 769 ISGVYSGDAAVAGSNRAHQMTGCIQAGYIMSDNFN-----HVPRLRFPAGGDSIRGY- 822
Db 949 ASA-----LFTSTPTMINSGLNTGSIYVYV--R--GTMAGVQAYN 987
QY 823 AHDSLSPIISKGLYTG--OVLAVGTAEXYNEFPKDLRLAVFGDIGNAVKGFTNDTKI 879
Db 988 QYDS-----YTNAGRDQOIRIAGTAE-----ACGAPADVOF 1017
QY 880 G-----AGVGVRRASPYGOVRVDATGVK 903
Db 1018 GNSWVRADVG--ASPIGFPITVNLGVE 1043

RESULT 9
YDF6_SCHPO
ID YDF6_SCHPO STANDARD; PRT; 475 AA.
AC 010478;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C17C9.06 In chromosome I.
GN SPAC17C9.06
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21648401; Pubmed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham S., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goole A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

```

```

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellion J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Filtz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wambutt R., Punelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Jelaune V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerretti L., Lowe T., MacCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RA "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -I- SIMILARITY: BELONGS TO THE UPE0140 (CGI-51) FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: Z73099; CAA97352.1; -.
CC DR InterPro: IPR000184; Bac_surfng_D15.
CC DR Pfam: PF011103; Bac_surface_Ag; 1.
CC DR Hypothetical protein.
CC KW SEQUENCE 475 AA; 51762 MW; 8FEC5B5F4B5DE1S CRC64;

Query Match 3.4%; Score 159; DB 1; Length 475;
Best Local Similarity 20.8%; Pred. No. 0.1; Mismatches 210; Indels 136; Gaps 25;
Matches 113; Conservative 83;

QY 442 FEQSSSRTEPAQVDESTLEPIETVELDGIIMDISPIEFASNLIDKLNIVAKARH 501
Db 5 FSTSFSPDIPAVNESKIS-AETFKSLEILANSTLPVGISSI----- 49
QY 502 LYDMDDRYLAINHD-DGVNSILGRISD-AVSARAILPDESENEYIDLPERTALNR 559
Db 50 -----RVTGAHHTRPSPFIRKVLKTCIDTSKPARSRSL--ETLNAIOETTGNIAMNF- 99
QY 560 KTPADVYQSKKVPV-----YVFVASKPPRGQIGLGMGSDTGTRLVTKFEH 605
Db 100 -----VYETANIKIDIRASSVSQDDDLDTVTOYKRPKPR--LYVETGIDVG-NVSGNVHA 150
QY 606 NLINRGYQAGAEIRLSEDKKGVKLATKPLSH--PLNDQLRATLGYQGEVFGHS----- 658
Db 151 NVLARNFV-GGAELLISGVNYSYTRNRSYMSVFEPRVADPTRLRFN-----GHSNLRDN 205
QY 659 -----TNGFPLSTR-----TLEHISRSIIQNGMWNR-TYSLKRLDKLTKQAPE 703
Db 206 KSISSHDLITRGTITSLQHDIMWSEHLLSQNL--WROVTHLTERAASPSVRLEADS 261
QY 704 TWODLPVDFVNGKPSQEAALLAGVAVHKTVADNLVPMRGYRQRTSLEYSGSG-LVSDANM 762
Db 262 LKQSLSTYIT-----RPTRHIMPTRGDGVYRQTLBELAGFLPGDASF 305
QY 763 AIAIRAGISGVYSGDNNAVGSNRAHQMTGIGIA--GYISDNFNHYPRLPFPAGGDSIR 820
Db 306 LKSE-----FWGQKAAVALNSSRSVSLSLSARIGALHSLNKQVSLDPRFMLGSGSTLR 358
QY 821 GYANDSLSPIDSKYILNGGQVLAIGTAY-----NTEFPKDLRLAVFGDIGNAVKGFTND 876
Db 359 GFSEDRIGPKDGRSLGTATYAFMSLSLFPKPVYDASKPRPDLFAVAG-----GLSNI 413
QY 877 TK-----IGAGVGRMASPYGOVRVD-----VATGQKEGNGNFKILHFFIGT 917
Db 414 TSPNPGGYKSKPCLISTGLGVYANPAPARFELNFTLPATTEKDIGR-GLOFGAGI 472

```

OY 918 PF 919  
DB 473 DF 474

RESULT 10  
SCA4\_RICAU STANDARD; PRT; 991 AA.  
AC 09A964;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)  
DE (Protein PS 120) (Fragment).  
GN SCA4 OR D.  
OS Rickettsia australis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=787;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sekeyova Z., Roux V., Raoult D.;  
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the  
RT 'gene D' coding for an intracytoplasmic protein."  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; AF187982; AAK31302.1; -.  
KW Antigen.  
FT NON\_TER 1 1  
FT SEQUENCE 991 AA; 108518 MW; F2ZDB94DDC09749A CRC64;  
SQ

Query Match 3.2%; Score 150.5; DB 1; Length 991;  
Best Local Similarity 19.0%; Pred. No. 0.91;  
Matches 186; Conservative 136; Mismatches 335; Indels 321; Gaps 45;

OY 45 HDTAINOK-----AGNPVYLTPEDIOARLNAAGLNAPQSOAL-DVNFDDQS--PISR 97  
DB 7 YESGIDEKKOEKAAALQPTLDADGFSFTPASSTOSTPSISTLSDTLSHDQTSPIRK 66  
OY 98 -----IGESPPRLGLDMSVIEETPRLSLELFAQESTEM--GIRPNDYIPRY 142  
DB 67 AVRETIIOPOKDELAEQ-----IHKDLAALVDRDLAEOKREIEEKKKNTLSAF 117  
OY 143 QGEOPNSEVVPPTLEPKPGLIKRL-----YARLFNDGVNKKV----- 180  
DB 118 FGNPANRELIDKAL--EKPELKKLEAIETAGLKNVFLTYIAANGYSGGFPMOMENI 174  
OY 181 --PRLKAKFYOSSGSESAIGSSHQKTEPYANITAALEDITQESAMDLNGSIPRLKOTA 238  
DB 175 SASDLRAIVVNDAGDELCTINETTKTKPFT--VAKODQO--VOIN----- 218  
OY 239 LVAARAVGYVDIDLSIINSIGEVDIITHDGEPIYIDYRAVEVGEG--ADDKA--FT 293  
DB 219 -----SYRIDPPI--KDKKADGSMH-----LSMALALADTKPSKDAVYFT 259  
OY 294 TVADEVPLLLIGDVFHGKYEYTKNLLENASAEHGYPDGRWLDRSVDVLLPNTADVSLIY 353  
DB 260 AHYEGR-----NGRPOLK-----EISSPKPLKFAG-----DQPRVAATIE 295  
OY 354 DTGYQYRFDEVEFTTIDPKTQQLTTPDKPLVYKRELLEQOLLTVNKG--AVNLQAVRALS 411  
DB 296 HGGRYI-----TLAVTRGKYKEMKKEVELHOGOSVDLSQRTIARDLTKVQGRS 342

OY 412 NDLIATRFNMVNTETVEPEREDIQNDQVSFEQSSSRTERPAQVDESTLEPIYETVELTD 471  
DB 343 QETL-----QPIITPMOEL-----KSIETPTTQVPPITPDSPLO--TETAAQMPQ 387  
OY 472 GILMDISPIEFESA-----SNLIODKLNVLAAKARHLXMDPDRLVLAIHNDGVNSIIGRI 527  
DB 388 S--QOVNPNLLNATALSSGMODLLNVNA-----GLTKKEDKNGTQI----- 427  
OY 528 SDAVSAVARAILPDESENE--VIDLPETALANRKP-----ADV 565  
DB 428 -DLINNAATGAILNNEKEKQANIIVLTENTVNNALFPDTKVAGVNAVLENIKNNQNTPD 486  
OY 566 YOSKAPLYVFNVSOK-----PRDQ-----IGLWGSDDTGRVY-----TKRENH 607  
DB 487 EKSKMLEATVALNSENENAPKOKQOLEKAVDVGMLKDDTSRAVALIDITDVIKSNL 546  
OY 608 INRDGYOAGAEHLRSEDKKGKVLATKPLSHPLNDQLRATIGYQOEVFHSGTNGFDLSTR 667  
DB 547 STRD-----KOTMLIAV-----GDKVASELSMA 570  
OY 668 TLEHEISRTIIONGWNRRYS-----LRYRLDKLTQAPPEWODLPVDFVNG--KPS 718  
DB 571 E-KOOLLGSVLLKGVETQVLSPEQOOLMOONLDKITA--EQTNAKITEVGGILANPA 625  
OY 719 QEALLAGVAVHKTVADNLVNPNGYRGORYSLEVGSSGLVSDAMALARAGISG-----Y 772  
DB 626 FNTIAKTEAIQNTTVKVLDSPIKAEIKGETLE-----STIKVVAESPLNDQKVDIV 677  
OY 773 YSRGDNAYSNRHAKMTGTGIOAGYIWSDNFNHVPYRLRFPFAGDQSIGVAHDSPLSD 832  
DB 678 KKGGE-AIASHKTKMSTERTISA-----IESVEGVASITALED 715  
OY 833 KGYLVGGVLAAGTAEVNYEPMKDLRLAVFGDIGNAYDKGTNDRTKIGAGVGRMASPVG 892  
DB 716 KRLMTKGLVDGIYEDKANPEMTKAVSRGVDKSTARREDQALMDA-----AS 762  
OY 893 QVRVDVAT-----GVKKE 905  
DB 763 EVALDRETQNTFKGLKEQ 780

RESULT 11  
SCA4\_RICPR STANDARD; PRT; 1022 AA.  
AC 09ZD49; 09ZD48; 09AJ36;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)  
DE (Protein PS 120).  
GN SCA4 OR RP498/RP499.  
OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=782;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=Magrid E;  
RX MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RT mitochondria."  
RL Nature 396:133-140(1998).  
RN [2]  
RP SEQUENCE OF 11-1016 FROM N.A.  
RA Sekeyova Z., Roux V., Raoult D.;  
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the  
RT 'gene D' coding for an intracytoplasmic protein."  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

```

CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 234.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ235272; CAAL14951.1; ALT_FRAME.
DR EMBL: AJ235272; CAAL14950.1; ALT_FRAME.
DR EMBL: AF200340; AAK31305.1; -.
KW Antigen; Complete proteome.
FT CONFLICT 11 15 EEPDL -> RPLV (IN REF. 2).
FT CONFLICT 365 365 H -> Y (IN REF. 2).
FT CONFLICT 413 413 MISSING (IN REF. 2).
FT CONFLICT 957 957 G -> R (IN REF. 2).
SQ SEQUENCE 1022 AA; 114410 MW; 03230E3A663A9622 CRC64;

Query Match 3.2%; Score 149; DB 1; Length 1022;
Best Local Similarity 20.7%; Pred. No. 1.2;
Matches 178; Conservative 131; Mismatches 303; Indels 248; Gaps 45;

QY 74 GLNAKRP-----SQALDVYVFDQSP--ISRIG-----OSPP 104
DB 267 GPNGKDPQKLTSSPQPKFVGTDGDAVAIEHGGEITLAVTRCKYEMKEVALNHQGS 326
QY 105 LGDMSVYEETPTSLSELAEQESTEMGINPNDIPEYQGBOPSEVVPPTLEPERGL 164
DB 327 VALSQTAEELTVHGGP---SHEHKKIITIPNQL-ESSIEQHNSQOVPIT----- 374
QY 165 IKRLVARENDGVNKPRLAKFYOSSQSETSAIGSSHOKTEPYANIKALEDTGESA 224
DB 375 -----TFNKSQD--PKISQIHQLPQAQSSGI-----PNVLNAANALSTSMODLL 419
QY 225 MDLNGSIPR---LRQPLVAARAVG-----YDIDLSTIRNSIGEVDT-- 265
DB 420 NNINSLVTKMODINKOSDLKEAIALNNKSKDPAEKQYNITDLA--KNFSKKDITAD 477
QY 266 ----IHDLGEPYIDYRAVEVREG--ADDKAFTVADEVPPLIGVFHNGKYEKKNL 318
DB 478 AKVNVAVNTLETIONDQNTLDIKKSKILEDTVATLNSENEL-----KQKQOI 526
QY 319 IENASAEHGFQDGRWLDKRSVDVILPNTADVSLITDTGTQYRPEVVEFTIDPTNQLTT 378
DB 527 LE-----KVVDI-----GLSTKDDISRVAVADSINDTYI--KSNITANE 562
QY 379 DPKLPVKRELLEQLLVNNGEAVNLQAVRALSNDLATRFVNNVNTIEVPEREQION- 437
DB 563 DKRKIFI--TVFOQ---INSYFSSNAKKULLBSILKKT---AEQVLSPEQOOLMNO 612
QY 438 --DOVSFEQSSSRTEPAQVDESTLEPIETVELTQGI-----LMDISPIEFS-ASNLI 488
DB 613 MLDNITTEHTKRPDTE--KVNNTLLEPLSNTALKTNIQWTSNVLD-SPVQIEMKSKLI 669
QY 489 QOKLNIYAAKARLLYMPDORVLAINHDDGVNSIIGRISDA-----VSAVARAI 538
DB 670 QVVTKTVAESA--LVE-PKDKTEIVK--GIGKTVIH-SDTSLPLHDKVYIMGSVAKGI 722
QY 539 LPDESENEVID-----LPERT-----ALANRKTPAADVOSKVPV--VYFVA 578
DB 723 V--ESKNDLLDRLLIAGLVDSIYEAKGNAAVYHAISMSIANSINSSEKALEKRSDDV 780
QY 579 SDRPRGQOI-GLWGSDGTGRLVTKFEHNLINDGVOAGA-----ELRLSEDKKGVKL 630
DB 781 SEKVLDEKREIONDLREKAKONINESKLDHDIYKNKTODVAAKLKVITVLDNSGQRCVSE 840
QY 631 YATKPLSHPLND-----QLRATLIGY-----QEVFGHST-----NCFDLSTR 667
DB 841 EAPKKVSSLNDISKRTIEKINNLRAVLSDQGNLKTFFEEKKDEATKVKVDELVAFAFNKSS 900

```

```

QY 668 TLEHE--ISRSIIQGNKRYSLRYLDDKLTQAP---PEFMQDLPVDFNCKPSQEAALL 723
DB 901 TEEQONFKISNLDINKTSLRSREIRLQIIDLNLKAOAKRAETENI-----SAKTEADVAVI 955
QY 724 AGVAVHKVAVANLVNPMGYSRORSLEYSSGSLVSDAMALAIACISGVISGDAAYASN 783
DB 956 SGSKSLKPSID-----EPYQKAMVYERDVRDIDKIKIMSALI-----NARDS- 1001
QY 784 RAHOMTGGIAGYIMSDNFN 803
DB 1002 -----IQSENFN 1008

RESULT 12
ACF7_HUMAN STANDARD; PRT; 5430 AA.
ID ACF7_HUMAN
AC Q9UPN3; Q9UPK0; Q9UG9; Q9H540; Q75053;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Actin cross-linking family protein 7 (Macrophin) (Trabeculin-alpha)
DE (620 kDa actin-binding protein) (ABP620).
GN ACF7 OR ABP620 OR KIAA0465 OR KIAA1251.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20001959; PubMed=10529403;
RA Okuda T., Matsuda S., Nakatsugawa S., Ichiotani Y., Iwashashi N.,
RA Takahashi M., Ishigaki T., Hamaguchi M.;
RT "Molecular cloning of macrophin, a human homologue of Drosophila
RT kakapo with a close structural similarity to plectin and dystrophin.";
RL Biochem. Biophys. Res. Commun. 264:568-574(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20026884; PubMed=10559237;
RA Sun Y., Zhang J., Kraeft S.-K., Auclair D., Chang M.-S., Liu Y.,
RA Sutherland R., Saigia R., Griffin J.D., Ferland L.H., Chen L.B.;
RT "Molecular cloning and characterization of human trabeculin-alpha, a
RT giant protein defining a new family of actin-binding proteins.";
RL J. Biol. Chem. 274:33522-33530(1999).
RN [3]
RP SEQUENCE OF 868-2350 FROM N.A.
RX TISSUE=Brain;
RC MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirosewa M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
RN [4]
RP SEQUENCE OF 1544-5057 FROM N.A.
RA Corby N.;
RT Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE OF 3734-5430 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98116662; PubMed=9455484;
RA Seki N., Ohira M., Nagase T., Ishikawa K.-I., Miyajima N.,
RA Nakajima D., Nomura N., Ohara O.;
RT "Characterization of cDNA clones in size-fractionated cDNA libraries
RT from human brain.";
RL DNA Res. 4:345-349(1997).
CC -1- FUNCTION: F-ACTIN-BINDING PROTEIN WHICH MAY PLAY A ROLE IN CROSS-
CC LINKING ACTIN TO OTHER CYTOSKELETAL PROTEINS. ALSO BINDS TO
CC MICROTUBULES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.
CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.

```



CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 37 SPECTRIN REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB029290, BAA83821.1; -  
 DR EMBL: AF141968, AAF06360.1; -  
 DR EMBL: AB033077, BAA86565.1; -  
 DR EMBL: AL137853, CAC15920.1; -  
 DR EMBL: AB007934, BAA32310.1; -  
 DR HSSP: 001082, 1BR.  
 DR -----  
 DR Genes: HGNC:13664; MACF1.  
 DR InterPro: IPR001589; Actbind\_actu1n.  
 DR InterPro: IPR001715; Calponin-like.  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR003108; GAS2.  
 DR InterPro: IPR001452; SH3.  
 DR InterPro: IPR002017; SPECTRIN.  
 DR Pfam: PF000307; CH; 2.  
 DR Pfam: PF00435; spectrin; 36.  
 DR Pfam: PF02187; GAS2; 1.  
 DR ProDom: PD000012, EF-hand; 1.  
 DR SMART: SM00033; CH; 2.  
 DR SMART: SM00054; EFh; 2.  
 DR SMART: SM00243; GAS2; 1.  
 DR SMART: SM00150; SPEC; 35.  
 DR PROSITE: PS00019; ACTININ\_2; 1.  
 DR PROSITE: PS00020; ACTININ\_2; FALSE\_NEG.  
 DR PROSITE: PS00021; CH; 2.  
 DR PROSITE: PS00018; EF\_HAND; 2.  
 DR PROSITE: PS00002; SH3; FALSE\_NEG.  
 DR Actin-binding; Cytoskeleton; Calcium-binding; Repeat; SH3 domain.  
 FT DOMAIN 1 295  
 FT 1 76 181  
 FT DOMAIN 194 295  
 FT REPEAT 314 355  
 FT REPEAT 591 623  
 FT REPEAT 680 784  
 FT REPEAT 786 800  
 FT DOMAIN 871 923  
 FT REPEAT 1250 1272  
 FT REPEAT 1287 1342  
 FT REPEAT 1455 1534  
 FT REPEAT 1547 1659  
 FT REPEAT 1815 1891  
 FT REPEAT 1932 2042  
 FT REPEAT 2260 2280  
 FT REPEAT 2372 2395  
 FT REPEAT 2510 2507  
 FT REPEAT 2621 2728  
 FT REPEAT 2731 2838  
 FT REPEAT 2841 2945  
 FT REPEAT 2987 3024  
 FT REPEAT 3136 3163  
 FT REPEAT 3187 3274  
 FT REPEAT 3277 3383  
 FT REPEAT 3386 3492  
 FT REPEAT 3495 3601  
 FT REPEAT 3604 3673  
 FT REPEAT 3713 3819  
 FT REPEAT 3832 3927  
 FT REPEAT 3982 4043  
 FT REPEAT 4046 4152  
 FT REPEAT 4155 4262  
 FT SPECTRIN 29.

FT REPEAT 4265 4371 SPECTRIN 30.  
 FT REPEAT 4374 4481 SPECTRIN 31.  
 FT REPEAT 4484 4590 SPECTRIN 32.  
 FT REPEAT 4593 4700 SPECTRIN 33.  
 FT REPEAT 4707 4808 SPECTRIN 34.  
 FT REPEAT 4811 4917 SPECTRIN 35.  
 FT REPEAT 4920 4985 SPECTRIN 36.  
 FT CA\_BIND 5030 5054 SPECTRIN 37.  
 FT CA\_BIND 5096 5107 EF-HAND 1 (POTENTIAL).  
 FT CA\_BIND 5132 5143 EF-HAND 2 (POTENTIAL).  
 FT DOMAIN 5276 5283 POLY-SER.  
 FT DOMAIN 5355 5370 4 X 4 AA TANDEN REPEATS OF [GS]-S-R-[AR].  
 FT CONFLICT 1 72 MSSDETLESCRSRCSRSRSYSRSRSGSLSPCPDP  
 FT TLPWNLPLHOKRKSODSVLDPDAERAVRY -> MEPLYM  
 FT AGIPIRVSGLQPLPPEKFOCTASAVAVI (IN REF.  
 FT CONFLICT 2)  
 FT CONFLICT 575 594 VAISSEDEGNLRVEYELLS -> GPSAPLMKRAISDLCKN  
 FT YCL (IN REF. 1).  
 FT CONFLICT 1487 1487 A -> T (IN REF. 2 AND 3).  
 FT CONFLICT 1963 1963 V -> A (IN REF. 1).  
 FT CONFLICT 2052 2052 E -> D (IN REF. 1).  
 FT CONFLICT 2083 2083 E -> K (IN REF. 1).  
 FT CONFLICT 2290 2290 M -> V (IN REF. 2).  
 FT CONFLICT 2321 2321 C -> Y (IN REF. 1).  
 FT CONFLICT 2343 2363 MISSING (IN REF. 2).  
 FT CONFLICT 2344 2350 SLPSVG -> EXRLPKI (IN REF. 3).  
 FT CONFLICT 2523 2523 O -> R (IN REF. 2 AND 4).  
 FT CONFLICT 4670 4670 S -> T (IN REF. 2).  
 FT CONFLICT 4833 4833 MISSING (IN REF. 2).  
 SQ SEQUENCE 5430 AA; 620346 MW; 91ADB7F7580B440B CRC64;

Query Match 2.9%; Score 138.5; DB 1; Length 5430;  
 Best Local Similarity 17.9%; Pred. No. 56;

Matches 164; Conservative 137; Mismatches 288; Indels 327; Gaps 42;

QY 14 PVALAAYPLMTSQAQAQ-----NNPANI-INHYPAHDTAIINQAKG-NPPVL 60  
 DB 2873 PVQLES--SLRSKAMLEVEKRRSLLEILNSADILLINSEADEGIRBEKAGINQNM 2930  
 QY 61 LTPBQIDARLNA-----AGLNKQOSALDVNFDDSPISRIGESGPL 105  
 DB 2931 AVTEELQAKTSGLEEMTORLREPOESFKNIEKKGAKHOLEIFD-----ALGSA 2981  
 QY 106 GLDMSVIEETPTLSLELFOESTENGINPN-DYIPRY-OG----- 144  
 DB 2982 -----CSNNLEKLRHQEYVLALEPOVDLIRNFOTGLVEDADPGSDASQLHQAE 3032  
 QY 145 -----EOPNSEVYVVPPTLEPERGL-----IKRLVARTLN-----DGVNKVPRL 183  
 DB 3033 VAQGEFLEVAGQVNSGCVM--MENKLEIGGFHCRVRKREMFSQLADLDDELDMGALGR- 3088  
 QY 184 KAKFYOSSGGETSAIGSSHOKTEPYAN-----IKALED-----ITGESANDLNG-- 229  
 DB 3089 -----DSDLSQSIQIEDVRLFLNKIHVLKLDIEASEAEGRHMLEEGTLDLGLK 3137  
 QY 230 -----SIPRLQRTALVAARAVGY-----DI 250  
 DB 3138 RELBALNKQCGKFLTERKAKQBOLETLGRVEDPYRKLKGLNDATTAABEAALQWVVG 3197  
 QY 251 DLSIIRNSIGEVNIIHDLGEPYIDYRAVEVRGEGADDKA----- 291  
 DB 3198 EVELINQQLADFMFOKEQYDPLQMKIQVNGJGGLIQSAGDCVQGLHDMEEFINAR 3257  
 QY 292 FTT----VADEVPLLIGDVFHGKYE-----TKNLIEN--ASAEGYEDGRW 333  
 DB 3258 WNTLNKKVAQRIQLOEALHLCGFQDALEPLLSWLADTELLANQKPPSAEYKVVAAQI 3317  
 QY 334 LDSS-VDVILPNTADYSLIYDGTQYRDEYVFTIDPKTNDLTDPKLPVKRLELE- 391  
 DB 3318 QEBKLLQRLDDKATVDYMDIAEGGR-----AQSAEL-ADREKITQGLSLES 3365  
 QY 392 ---QLITVNGEAYNLQAVALS-----NDLIATRYFMNVNTEIVFPERQOI-- 436  
 DB 392 ---QLITVNGEAYNLQAVALS-----NDLIATRYFMNVNTEIVFPERQOI-- 436

```
Db 3366 RWTETLSKAAAROKOEDILVLAKOFHETAEPIISDFLSTYEKKLANSEFVGTOTAKIQOO 3425
Oy 437 -----NDQVSFEQSSSSRTEPAOVDESTLEPIETV-----ELTDGI 473
Db 3426 IIRHKALEEDIENHATDVQAQVIGOSLSLTSIPA--EOGVISEKIDSLQARYSETIODRC 3483
Oy 474 LMDISPIEFSANL-----IODKLNLVAAKARHLYDMEDDRLAINHMD 517
Db 3484 CKKALLDQALSNARLFGDEVEVLNWLAEVEDKLSVYK-----DEFOD-VLHRQHND 3537
Oy 518 --GVNRSILGRISDAVASV--ARAILPDESENEVIDLPER--TALANRKTADVYOSKKVP 572
Db 3538 HLALNEIYVNRKKNVQAIKNQOALLKQTTGHEVLLIQEKLQGIKTRVADITVTSKALR 3597
Oy 573 LYVFAASDKPRGQIGLGMGSDTGTRLYTKFEHNLINRGQYQAGAEIRLSEDKKVKLYA 632
Db 3598 TL-----DQARQLARKFQSTYEELTGW-----LREVEEE-----LA 3628
Oy 633 TKPLSHPLNDQRLATLGYQOEVEFGHSTNGFDLSTRTLEH-----EISRSIIONGMN 684
Db 3629 TSGGSGPTGEGQIPQFOQORKE-----LKEVMEHRLVLDIVNEVSRALLLVPR 3678
Oy 685 RTYSRLYRLDKLKTQA 700
Db 3679 ---AREGLDKLVSDA 3690

RESULT 13
G6PD_STRPN STANDARD: PRT: 495 AA.
ID G6PD_STRPN 054537;
AC 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD).
GN ZMF OR SP1243.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RA Tetteelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen D.,
RA McDonald L.A., Feldblum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
RN [2]
RP SEQUENCE OF 281-433 FROM N.A.
RC STRAIN=ATCC 6323;
RX MEDLINE=98125733; Pubmed=9466257;
RA Coffey T.J., Enright M.C., Daniels M., Morona J.K., Morona R.,
RA Hyman-Lewis W., Paton J.C., Spratt B.G.;
RT "Recombination exchanges at the capsular polysaccharide biosynthetic
RT locus lead to frequent serotype changes among natural isolates of
RT Streptococcus pneumoniae.";
RL Mol. Microbiol. 27:73-83(1998).
RN [3]
RP SEQUENCE OF 281-433 FROM N.A.
RX MEDLINE=98192409; Pubmed=9533725;
RA Coffey T.J., Enright M.C., Daniels M., Wilkinson P., Berron S.,
RA Renoll A., Spratt B.G.;
RT "Serotype 19A variants of the Spanish serotype 23F multiresistant
RT clone of Streptococcus pneumoniae.";
RL Microb. Drug Resist. 4:51-55(1998).
```

```
CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-
CC 1,5-lactone 6-phosphate + NADPH.
CC -1- PATHWAY: Pentose phosphate pathway; first step.
CC -1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
```



```
CC ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR
CC RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
CC CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 6 ACYL CARRIER DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF004835; AAC45930.1; -.
DR HSSP: P14687; 1AMU.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR001242; Condensatn.
DR InterPro: IPR003880; Pantne_attach.
DR InterPro: IPR003799; Ser_estrs_site.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00501; AMP-binding; 6.
DR Pfam: PF00550; pp-binding; 6.
DR Pfam: PF00668; Condensation; 6.
DR Pfam: PF00975; Thioesterase; 1.
DR PRINTS: PRO0154; AMPBINDING.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 6.
DR PROSITE: PS00455; AMP_BINDING; 6.
DR PROSITE: PS50075; ACP_DOMAIN; 6.
KW Ligase: Antibiotic biosynthesis; Phosphopantetheine;
KW Multifunctional enzyme: Repeat.
FT REPEAT 466 1038 DOMAIN 1 (ASPARAGINE-ACTIVATING).
FT REPEAT 1521 2070 DOMAIN 2 (GLUTAMINE-ACTIVATING).
FT REPEAT 2536 3113 DOMAIN 3 (TYROSINE-ACTIVATING).
FT REPEAT 3590 4149 DOMAIN 4 (VALINE-ACTIVATING).
FT REPEAT 4606 5203 DOMAIN 5 (ORNITHINE-ACTIVATING).
FT REPEAT 5658 6245 DOMAIN 6 (LEUCINE-ACTIVATING).
FT DOMAIN 970 1037 ACYL CARRIER (ACP) 1.
FT DOMAIN 2007 2074 ACYL CARRIER (ACP) 2.
FT DOMAIN 3045 3112 ACYL CARRIER (ACP) 3.
FT DOMAIN 4080 4147 ACYL CARRIER (ACP) 4.
FT DOMAIN 5124 5191 ACYL CARRIER (ACP) 5.
FT DOMAIN 6167 6234 ACYL CARRIER (ACP) 6.
FT BINDING 1000 1000 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 2037 2037 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 3075 3075 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 4110 4110 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 5154 5154 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 6197 6197 PHOSPHOPANTHETHEINE (BY SIMILARITY).
SQ SEQUENCE 6486 AA: 724011 MW: 4934900A F07DE786 CRC64;

Query Match 2.98; Score 136; DB 1; Length 6486;
Best Local Similarity 18.5%; Pred. No. 1e+02;
Matches 139; Conservative 102; Mismatches 285; Indels 226; Gaps 33;

OY 44 AHDPAINQAKGNPNVLTPEQIOARLNAAGL--NAKP-----OSOA 83
| | | | | | | | | | | | | | | | | | | | | |
Db 269 ADDVVFSGSVSGRPP-----QIOGIESMVGFLTINTPTRVQTNKOOTFSELLQTVQKOA 322
| | | | | | | | | | | | | | | | | | | | | |
OY 84 LDVNVFP-----DOSPIRIGSGSPGLG----- 107
| | | | | | | | | | | | | | | | | | | | | |
Db 323 LASATYFAPFLYEIQSTFTVLKQELIDHLVTFENYPDHSMKMLEESLGQTFVESGDEQTS 382
| | | | | | | | | | | | | | | | | | | | | |
OY 108 -DMSVIETTPPLSELEFAOSTEMGINPNDYIEPYGEOGRN--SEVVVPPLEPERKGL 164
| | | | | | | | | | | | | | | | | | | | | |
Db 383 YDLNVVVALASN--ELYVKLSYNAAYESSFVNRIECHLRTVIDQVIGNPHVHLHEIGI 440
| | | | | | | | | | | | | | | | | | | | | |
OY 165 I-----KRLYARLFNDGNKVPRLAKAFYQSSQSGTSAIGSHKTEPYANIKALEIT 220
| | | | | | | | | | | | | | | | | | | | | |
Db 441 ITEEEKQQLLVAYMDTAETAYPRDKTIF-----ELIAQASRTPAKKAVCGEDTLT 491
| | | | | | | | | | | | | | | | | | | | | |
OY 221 QESAMDLNGS IPR-LRQTAIVARA VGY--YDIDL-----SIIRNSIGEVVITIHDLGE 271
| | | | | | | | | | | | | | | | | | | | | |
```

```
Db 492 YQELMERASQULANALREGGIASGTSIVMAEHSLELYAIMAVLSSGAAYLP-----D 545
| | | | | | | | | | | | | | | | | | | | | |
OY 272 PYIDYIRAVEVYEGADAKFAFTVADEVP-----LIGVFHKGKYEFTKKNLEINASAEH 326
| | | | | | | | | | | | | | | | | | | | | |
Db 546 PEYPODRIRQYL-----LDDSQTLTLTQSHLPQINIFAGSVL-----YLDRLSLEYEGST- 596
| | | | | | | | | | | | | | | | | | | | | |
OY 327 GYEDGRWLDRSVYVILPONTAD--VSLIYDNGTOYRFDEVEFTIDPKTNQTLTPDPKLP 384
| | | | | | | | | | | | | | | | | | | | | |
Db 597 -----FAPESKPPDLAYMYITSGS-----TGNPKGAMIT----- 625
| | | | | | | | | | | | | | | | | | | | | |
OY 365 VKRELLLEDQLITVY-----MGEAYNLQAVRALNSMDLATRFENMV-----NTEIYPEREON 437
| | | | | | | | | | | | | | | | | | | | | |
Db 626 -HQGLVNTYIMANKKYVOGEAVDPPLYSISFDLVTSIFPFLSGNTIHYVRADKVO- 683
| | | | | | | | | | | | | | | | | | | | | |
OY 438 DQVSFEQSSSRTERPAQVDESTLEFVIEVELTDLGILMDISPIEFSASNL--IDOKLNL 494
| | | | | | | | | | | | | | | | | | | | | |
Db 684 -----VILDIIDKQKVGIIKLTPIPHKLEIHIDDKASSIRRFIVGENTL 727
| | | | | | | | | | | | | | | | | | | | | |
OY 495 VAAKARHLXDMFPDDRVLAINHDDGVNRSILGRISDAVSAVARALLPDESENEVIDLPERT 554
| | | | | | | | | | | | | | | | | | | | | |
Db 728 PTKLAKQIYDHFGENVQJFNE-----YGPTEYVVGCMIVLYDPOTTQDSVP- 775
| | | | | | | | | | | | | | | | | | | | | |
OY 555 ALANKRTPADVYQSKKVPPLYVPVASDKPRD-GQIGLGMSDPTGFRLYVFEHNLINRQY 613
| | | | | | | | | | | | | | | | | | | | | |
Db 776 -----GVPRAD-----NVQLYLDASMKQVPVGSGL-----EMTYAGDGV 809
| | | | | | | | | | | | | | | | | | | | | |
OY 614 QAGAEIR--LSEDK-----KGVKLYATKPLSHPLND--QLRLTGLYQGEVFGSHSTNG 661
| | | | | | | | | | | | | | | | | | | | | |
Db 810 AKGYRNRELTKEKFTIDNPFRTGKMYRTGLAKMLPDGMMEYAGMDYQVKIRGHRIEM 869
| | | | | | | | | | | | | | | | | | | | | |
OY 662 FDLSTRITLEHESRSII-----QNGGMNRTYS 688
| | | | | | | | | | | | | | | | | | | | | |
Db 870 GEITRILTQHEAVKKAAYIVKEDSGQNVLYA 901
| | | | | | | | | | | | | | | | | | | | | |

RESULT 16
ID BCSC_PSEFL STANDARD: PRT: 1279 AA.
AC PS8937;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose synthase operon protein C precursor.
GN BCSC OR WSGE.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SRW25;
RX MEDLINE=22013850; PubMed=12019221;
RA Spiers A.J., Kahn S.G., Bohannon J., Traviano M., Rainey P.B.;
RT "Adaptive divergence in experimental populations of Pseudomonas
fluorescens. I. Genetic and phenotypic bases of wrlnky spreader
fitness".
RL Genetics 161:33-46(2002).
CC -1- FUNCTION: Required for maximal bacterial cellulose synthesis (By
similarity).
CC -1- PATHWAY: Bacterial cellulose biosynthesis.
CC -1- SIMILARITY: CONTAINS 10 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AY074776; AAL71845.1; -.
KW Cellulose biosynthesis; Repeat; TPR repeat; signal.
```

FT SIGNAL 1 21 POTENTIAL  
 CHAIN 22 1279 CELLULOSE SYNTHASE OPERON PROTEIN C.  
 REPEAT 27 60 TPR 1.  
 REPEAT 62 94 TPR 2.  
 REPEAT 218 250 TPR 3.  
 REPEAT 306 339 TPR 4.  
 REPEAT 384 417 TPR 5.  
 REPEAT 460 493 TPR 6.  
 REPEAT 495 527 TPR 7.  
 REPEAT 606 639 TPR 8.  
 REPEAT 719 752 TPR 9.  
 REPEAT 787 820 TPR 10.  
 SEQUENCE 1279 AA: 138664 MM: EAYE42BE7810D201 CRC64;

Query Match 2.8%; Score 134.5; DB 1; Length 1279;  
 Best Local Similarity 19.1%; Pred. No. 10;  
 Matches 199; Conservative 140; Mismatches 334; Indels 369; Gaps 54;

QY 28 ALAQONNP-----ANITNHPADHTAINQAKGNPVYL-----TPEDIQARLNAGLNAK 78  
 DB 374 AQAERLNPGPGGAIALAGFOADNQDFDAEAGYRKYLARHPGDDALSLGINVLSQSGQ 433  
 QY 79 POSOALDVNFDDQSPISRIGESPPGLDMVIEETTPLSLEELFQESTEMGINNDY 138  
 DB 434 P-DEALKLI--DVSFPAQR-AKRAPSVKIN-----ALRAVQV-----K 468  
 QY 139 IPEYGEQPNSEVVPPTL--EPEKPG--LIKRLYARLFNDGVNKPRLKAFYQSSOS 193  
 DB 469 LARQGDGLKAAQAAVRYALADPENPWRPALARMYLR--DGIIRAKALIDGLKQSP 525  
 QY 194 GFTSAIGSSHQKTEPVANITALEDTQESAMDINGSIPRLQOTALVAARAVYDIDLS 253  
 DB 526 NQDVALYTS-----TLISAQJSEMKAQEA--TLGRIPTAQRTA-----DMN 564  
 QY 254 IIRNSIGENVVITHDGEPIYIDYRAVEVRGEGADKAFITVADDEVLLIGDVPHHKYE 313  
 DB 565 EL-----AIDIALHQOTD-----IAETARRGPRPALALLGRSEPLT-----RNKPE 607  
 QY 314 TKKNL-----TENASAEHGYPDGRMLDRSDVI--LPDNTADVSLIY----- 353  
 DB 608 RVAVLAAAVYEGAAQYG-----LDMNQVVENNPPTVDOKLLYANVLLKANKYSEAG 661  
 QY 354 -----DTGTOYRDEVEVFEFTIDPKTQOLTDPDKLPVKRELLBOLLTVNMGEA 401  
 DB 662 ELIREVOGQPLTFETGRQ-RYDDLILY----- 686  
 QY 402 YNLQAVRAL--SNDLIATREYNVNTETVPEREQIONDOVS-----PEQSSSSTEP 452  
 DB 687 YRKQADALREKNDLVAA--YDMLIS-----PALAQRPDALGVGALARMYAAASNGK-- 736  
 QY 453 AQVDESTLEPVIETVELTDGILMDISPIEFSASNLIDDKLNL--VAAKA--RHLYDMPDD 508  
 DB 737 -----KAMEL-----YAPLIOQNNNMARLOGLADIALKGDGRLAGQASD 777  
 QY 509 RVLAIHHDGVNRSI-----IGRISDAV-----SAVARAILPDE 542  
 DB 778 KALAL-----EFGNBEILTSARIYQGLKNSAEALLKALAIENAMAKAQVOAASAPGT 834  
 QY 543 SENEVIDLPERTALANRKPADYQSKKPVLYFVASDKRQDQIGIGMSDGTGLRYTK 602  
 DB 835 STNPFVGLP-----GQRROVTDLTVAGAVPPPI--DAPTKSVTSNAFASATSNLSDP 885  
 QY 603 F-----EHNLIINRDGYOA-GAELTSEDKGVK---LYATKPL 636  
 DB 886 FVPPSSIASIDSPELSPARALDTILDRGRGYVVGGLSVASNGGKLSITTVAEAFEA 945  
 QY 637 SHPLNDQ---LRAT-----LGY 650  
 DB 946 RHPVGVNTVALRPTPVHLSAGSVKASLSRFKGTGTEPAGSQSDSGVLAVAFENPDQGL 1005  
 QY 651 OOEVEGHSNTNGPDLSTRTLEHLSRSLIONGS--WNRTYSLRYRLDKLKTQAPPE----- 703  
 DB 1006 KADV--GVSPFLGLFYNTLVGCVSVSRPEFANSNFRYGANISRPRVTDTSVNSFAGSEDCAGN 1064

QY 704 TWODLPVDFVNGKPSOEALLAGV-----AVHKTVADNLVNPARGYRQRYSTLEVGS--GLV 757  
 DB 1065 KMGVTTANNGRGELSYDNQKIGVYGASLHLLGNVNDNR-----LELGGGIWYL 1117  
 QY 758 SDANMALIARAGISG-----YSEFGDNAVYSNRAHQTGTGIQAGYIWSDFNHVPY 807  
 DB 1118 RNNPRDTLTLGISGSAMTFKENODEFYTNGCYFESPOFFSILG--VPIRMAQSFDRFSY 1174  
 QY 808 RLRFAG-----GDOSIRGADHSLSPISSKGLTGTGGVLAATAEYNPEFM 854  
 DB 1175 QVKS SVGLQHTAGDADYFPEDSTLQATKN--NPKYDSTSKTG-----VG--YSFNAA 1223  
 QY 855 KDLRLA-----VFGDIG--NAYD 870  
 DB 1224 AEYLLSSRFYLGELTGLDMQD 1245

RESULT 17  
 P200\_MYCPN STANDARD; PRT; 1036 AA.  
 ID P200\_MYCPN  
 AC P75211: 050346;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein P200.  
 GN P200 OR MPN567 OR MP275.  
 OS Mycoplasma pneumoniae.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=2104;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 29342 / M129;  
 RX MEDLINE=96257197; PubMed=8675035.  
 RA Prof T., Hilbert H., Plagens H., Hermann R.;  
 RT "The P200 protein of Mycoplasma pneumoniae shows common features with  
 RL the cytoadherence-associated proteins HMW1 and HMW3.";  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 29342 / M129;  
 RX MEDLINE=97105885; PubMed=8948633;  
 RA Himmelfreisch R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,  
 RA Hermann R.;  
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
 RL pneumoniae.";  
 RN [3]  
 RP Nucleic Acids Res. 24:4420-4449(1996).  
 RN [3]  
 RP SEQUENCE OF 641-678 FROM N.A.  
 RC STRAIN-ATCC 29342 / M129;  
 RX MEDLINE=95075318; PubMed=7984111;  
 RA Prof T., Hermann R.;  
 RT "Identification and characterization of hitherto unknown Mycoplasma  
 RL pneumoniae proteins.";  
 RN [4]  
 RL Mol. Microbiol. 13:337-348(1994).  
 CC -1- FUNCTION: PROTEIN CYTOSKELETON-ASSOCIATED WHICH COULD BE AN  
 CC ACCESSORY STRUCTURAL COMPONENT IN CYTOADHERENCE.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC DR EMBL: U25989; AAC99815.1; -  
 CC DR EMBL: AE000027; AAB95923.1; -  
 CC DR EMBL: Z32646; CA83569.1; -  
 CC KW Cytoadherence; Structural protein; Repeat; Complete proteome.  
 FT DOMAIN 277 280 POLY-PHR.  
 FT DOMAIN 300 845 PRO-RICH.  
 FT DOMAIN 357 360 POLY-PHR.



```

Db 622 KQIGRPTFD-AEKVTIDLSKVYVQADASLNEKDIAVAANAISGRKADVTASTDLHLQ 680
QY 510 -----VLAINHDDGVNRSILGRISDAVSAYARALP-----DESENEVIDL----- 550
Db 681 NTVTAMKTNADSV-----VL-----AMGYKYLLVLPFVKNVKGDEPNTVOLTNGETV 732
QY 551 -----PERALANRK-TPADY-YQSKVPL-----YFVASOKPRD-GQIGLG 590
Db 733 TWTVINHVGSNSKDVKADKCTGVSLSLHDDIPLOTKIYEVKSSERPANYGCTEE 792
QY 591 WQSD-----TGRLVTKFEHNLIIRD-----GYOAGAEIR-----LSBDKKGVLYATKPL 636
Db 793 KGMNDVLDTHTRFTGKW-HAITNYDLKVGDKTLKAGTDSIAYILLEND-----NKDL 845
QY 637 SHPLNDOLRATLGYOQEVGHSINGFDLSTRLEHEISRSIIIONGWNRTSYLRYLDKL 696
Db 846 TETMNQALLAAL-----NEGSNKYKQAMSYYLEVER---ITGDEVNTQNTENYKELY 896
QY 697 KTOAPEPTMODLPVDFVNGKPSQOEALLAGVAHVAKTVADNLVNMKRGYRORYSLEVGSSGL 756
Db 897 RNTVTVTHTRPDP-----KPTK-----AVHNKKGEDI----- 923
QY 757 VSDANAIARAGISGVYSPGDNAVGSNRAHOMTGIOAGYIMSDNENHVPYRLRFFAGD 816
Db 924 -----NHGKYARGDVLSTEMTWDLK-GYDKFAEDTVD-----LATGV 960
QY 817 QSIRYANDLSLSPISD-----KGYLT-----GGQVLAVGTA 847
Db 961 SFDDYDETKVPIKDLIAVVKDSKSGEDITNOGTISWDKAGTYTISAKDPOAFTILAHGQ 1020
QY 848 EYNTEPMKDLRLAVGDDICN-AVDKFTNDTKIGAGVG-VRRNASPVGOYRVDAVATVKKE 905
Db 1021 ELRVTLPTVKVKNAGSDVYNLAEQNFQGRIKTNTVYNHPIKVPKPKDVYIKVGRQSON 1080
QY 906 GNPRL 911
Db 1081 GATIKL 1086

```

```

CC -1- SIMILARITY: CONTAINS 32 CADHERIN DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF231022; AAF61928.1; -
CC EMBL: AB011535; BAA32463.1; -
CC EMBL: AL157443; CAB75663.1; -
CC Genew: HGNC:3596; FAT2.
CC MIM: 604269; -.
CC HSSP: P00740; IEDM.
CC InterPro: IPR002126; Cadherin.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR001791; Laminin_G.
CC Pfam: PF00008; cadherin; 33.
CC Pfam: PF00008; EGF; 2.
CC PRINTS: PR00205; CADHERIN.
CC SMART: SM00181; EGF; 2.
CC SMART: SM00112; CA; 32.
CC SMART: SM00282; LamG; 1.
CC PROSITE: PS00232; CADHERIN_1; 14.
CC PROSITE: PS00268; CADHERIN_2; 32.
CC PROSITE: PS00222; EGF_1; 2.
CC PROSITE: PS01186; EGF_2; 1.
CC PROSITE: PS0025; LAM_G_DOMAIN; 1.
CC KW Transmembrane; Glycoprotein; Repeat; EGF-like domain; Signal.
CC SIGNAL 1 18
CC CHAIN 19 4349
CC DOMAIN 19 4048
CC TRANSMEM 4049 4069
CC DOMAIN 4070 4349
CC FT 34 148
CC DOMAIN 149 256
CC FT 363 458
CC DOMAIN 459 564
CC FT 565 669
CC DOMAIN 716 820
CC FT 821 925
CC DOMAIN 926 1032
CC FT 1033 1137
CC DOMAIN 1138 1242
CC FT 1243 1346
CC DOMAIN 1350 1448
CC FT 1449 1555
CC DOMAIN 1556 1660
CC FT 1661 1758
CC DOMAIN 1759 1872
CC FT 1872 1969
CC DOMAIN 1969 2070
CC FT 2071 2171
CC DOMAIN 2172 2272
CC FT 2273 2379
CC DOMAIN 2380 2481
CC FT 2482 2585
CC DOMAIN 2586 2691
CC FT 2691 2797
CC DOMAIN 2798 2906
CC FT 2907 3011
CC DOMAIN 3012 3113
CC FT 3114 3218
CC DOMAIN 3219 3321
CC FT 3321 3426
CC DOMAIN 3427 3531
CC FT 3532 3642
CC DOMAIN 3643 3773
CC FT 3773 3984
CC DOMAIN 3985 4022
CC EGF-LIKE 1.
CC EGF-LIKE 2.

```

RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

```
FT DOMAIN 3297 3302 POLY-SER.
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 459 459 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 655 655 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 789 789 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 966 966 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 1175 1175 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 1303 1303 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 1383 1383 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 1417 1417 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 1904 1904 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 1998 1998 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 2007 2007 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 2165 2165 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 2183 2183 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 2325 2325 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 2368 2368 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 2387 2387 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 2430 2430 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 2470 2470 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 2547 2547 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 2597 2597 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 2654 2654 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 3125 3125 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 3276 3276 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 3310 3310 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 3430 3430 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 3471 3471 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 3601 3601 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 3772 3772 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 3813 3813 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 3840 3840 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 3873 3873 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 3904 3904 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 3989 3989 N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT CARBOHYD 4117 4117 P -> L (IN REF. 2).
FT CONFLICT 4160 E -> G (IN REF. 2).
FT SEQUENCE 4349 AA: 479383 MW: 60500CA17AD398ED CRC64;

Query Match 2.88; Score 133.5; DB 1; Length 4349;
Best Local Similarity 17.48; Pred. No. 75;
Matches 199; Conservative 162; Mismatches 388; Indels 395; Gaps 50;

OY 52 AKAGNPVLLTPQO--IQARLNLNAGLNAKPOSOALDVNFDQSPISRIGQSP-GLID 108
DB 2145 ARDGGPTLSQSEEBVLTVNKNKSNPLFQSPYKRVRENTILTYPIHLHTQKSPFGLRLI 2204
OY 109 MSVIEE-----TTPLSLE---ELFQAQESTENGINPNDIPEYQEGOP 147
DB 2205 YNIVEEERPLMLFTTDFKTVGLTVGTPLDYESKTKHVFTRATDIALG-----SFESEA 2256
OY 148 NSEVVV-----PPTLE-----PEKPLIKRLVA-----RLFNDGV 177
DB 2257 TVEVLEEDVDNPNPTFSQVLYTTSISGLPAQVFPVLIQLASDDSGNRNDRVSYQIVEDGS 2316
OY 178 NKVPRLLAKFYQ--SQSGETSAIGSHQKTEPYANIKAALED-----ITQESAM-----D 226
DB 2317 D-----VSKRFQJNGSTQVQELDYEAQOHFHVKRAMDKDPPRLTGITLVVAVNSD 2371
OY 227 LNSIPRLKOTALVARANGVYIIDLSIINNSIGEVDVITHDGEPIYIDYRAVEVGE- 285
DB 2372 INDNPEFRPOPYEAVNS-----ELATCGHLVLYKQAIQIPDSRDTSRLELYLLSGNDR 2425
OY 286 -----GADKAFITVADEVLLGCV-----FH 308
DB 2426 HFTINSSGIIISFNLCKKLLDSSYNLRVGASDGVFRAT---FVYVNTTNANKYSPEFQ 2482
OY 309 HGKYEETKKNLIENASAEHGYFDGRWLDNRSDVLLPONTADVSLIYDTGCTQYRDEVEVFT 368
```

```
DB 2483 QHLYEAE--LAENAVGTRKVIDLLAIDKDSG---PYGTIDYTIINKLASEK-----FS 2530
OY 369 IDPKINOLTTTPDKLPVVRRELLEOLLYVMGEAYNLQAVRLASND----- 413
DB 2531 INP-NGQIAT-----LQKLDRENSTE--RVIAIKVARKGGGRVAFCTYKIILT 2576
OY 414 -----LIATRYENMNT-----EIVPEREQIONQVSEFQSSSRTEPAQVD 456
DB 2577 DENNDPQPFKASEYVYSIQSNVSKDSPYIQLVAYADQEGNADVITY-----SVNPEDLYKD 2632
OY 457 ESTLEPYIETVELTDCIL-MDISPIEFSA-----SNLQDKLNLVAAKAR----- 500
DB 2633 VLEINPVYGVVKKVKSJLQGLENOTLDFEIKADQDGPPIHNSLVPRLOVVPKVSLEPKRS 2692
OY 501 ---HLNDPMDRVLAINDDGVNRSILGRISDAVSAVARAILPDSESENVIDLPRTALA 557
DB 2693 EPLTFESAPED--LPEGSEIGIVKAVAAQ--DPVYISLVKGTTPESNKKQGVESLDPTGYI 2749
OY 558 NKRTPAD-----VYQ-----SKK 570
DB 2750 KYRKPMDESTRKLYQIDVMAHCLQNTDVSLVSNIQGVNDNRPVEADPYKAVLTEN 2809
OY 571 VPL-----YFVVASDKRRDQIGLGMSDPTGTRLVTFEHLINRD----- 611
DB 2810 MPVGTSVIQTVAIDKTDGRDQGVSYRLSADPGSNV---HELFAIDSESGMITTLQELD 2864
OY 612 -----GYQAGAEIRLS-----EDKKGVKL 630
DB 2865 CETQCTYHFHVAVYHGTQIQLSQALVOYSTTDENDNAPRASREYRGSVVENSEPEL 2924
OY 631 YAT-KPLSHPLNDQLRATLGYQOE-----VEGHSINGFD---LSTRTLBEHISRSIIONG 681
DB 2925 VATLKTLDADISEQNRQVTCYITTEGDPILGQFGISQVGEWIRISSRKTLDREHTAK----- 2979
OY 682 GMRNRYSLRYRLDKKQAPPEPMODLPVDFVNGKPSQSEALLAGVAKRYADNLVNMFR 741
DB 2980 -----YLLRVTASDQKFOA-SYVEIEIVLDVNDNSPQCSQLTYTKKHVEDV----- 3024
OY 742 GYRORYSLEVGSSGLVSDANMAIARAGISGVSEFSDNAYGNSRAHQMGIOAGYWS-- 799
DB 3025 -FPGHFILKVASATLDDPTGNAOI-----TYSL-----HGPO-AHEFKLPDPTGELTLT 3071
OY 800 -----DNFNHVPRLRLEFPFAGGDSIRGVAHDSLPSIDSKGYLTGQVLAVGTAEVYE 852
DB 3072 ALDRERKQVENLV---AKATDGGGRSCQADITLHVEDVNDNA-----PRF 3113
OY 853 FMKDLRLAVFGD--IGNAYDKGFTMDTKIGAGVGRMASP---VGQVAVDVATGVKEGN 907
DB 3114 FPSHCVAVFPNTYKTVAVAVVAFARDPDQGANAOVYVSLPDSABGHFSIDATGYIRLEK 3173
OY 908 PIKL 911
DB 3174 PLOV 3177

RESULT 20
TRX_DROME STANDARD; PRT; 3726 AA.
AC P20659; Q27255; Q27327;
DT 01-FEB-1991 (Rel. 17, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE TtHox protein.
GN TRX.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90192757; PubMed=2107543;
```



RA Mazo A.M., Huang D.-H., Mozer B.A., David I.B.;  
 RT "The trithorax gene, a trans-acting regulator of the bithorax complex  
 in Drosophila, encodes a protein with zinc-binding domains.";   
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2112-2116(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.  
 RA MEDLINE-95009521; PubMed-7924996;  
 RX Sedkov Y., Tiliib S., Mizrokhi L., Mazo A.;  
 RT "The bithorax complex is regulated by trithorax earlier during  
 Drosophila embryogenesis than is the Antennapedia complex, correlating  
 with a bithorax-like expression pattern of distinct early trithorax  
 transcripts.";   
 RL Development 120:1907-1917(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-Oregon-R;  
 RL MEDLINE-96100387; PubMed-8555104;  
 RA Tiliib S., Sedkov Y., Mizrokhi L., Mazo A.;  
 RT "Conservation of structure and expression of the trithorax gene  
 between Drosophila virilis and Drosophila melanogaster.";   
 RL Mech. Dev. 53:113-122(1995).  
 RN [4]  
 RP CHARACTERIZATION.  
 RX MEDLINE-95047388; PubMed-7958911;  
 RA Kuzin B., Tiliib S., Sedkov Y., Mizrokhi L., Mazo A.;  
 RT "The Drosophila trithorax gene encodes a chromosomal protein and  
 directly regulates the region-specific homeotic gene fork head.";   
 RL Genes Dev. 8:2478-2490(1994).  
 CC -1- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION  
 WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.  
 IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A  
 SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- MISCELLANEOUS: THIS PROTEIN HAS BEEN EXPERIMENTALLY SHOWN TO BIND  
 ZINC.  
 CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.  
 CC -1- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M31617; AAA29025.1; -  
 DR EMBL: Z50152; CAA90514.1; -  
 DR EMBL: Z50152; CAA90513.1; -  
 DR EMBL: Z31725; CAA83516.1; -  
 DR EMBL: Z31725; CAA83515.1; -  
 DR PIR: A35085; A35085.  
 DR HSSP: P20393; 1A6T.  
 DR TRANSFAC: T00850; -  
 DR Flybase: FBgn0003862; trx.  
 DR InterPro: IPR003889; FYrich.C.  
 DR InterPro: IPR003888; FYrich.N.  
 DR InterPro: IPR003616; PostSET.  
 DR InterPro: IPR001214; SET.  
 DR InterPro: IPR001965; ZnF\_PHD.  
 DR InterPro: IPR001841; ZnF\_Ling.  
 DR Pfam: PF00628; PHD; 3.  
 DR Pfam: PF00856; SET; 1.  
 DR SMART: SM00542; FYRC; 1.  
 DR SMART: SM00541; EYRN; 1.  
 DR SMART: SM00249; PHD; 4.  
 DR SMART: SM00508; PostSET; 1.  
 DR SMART: SM00184; RING; 3.  
 DR SMART: SM00317; SET; 1.  
 DR PROSITE: PS50280; SET; 1.  
 DR PROSITE: PS01359; ZF\_PHD\_1; 4.

DR PROSITE: PS50016; ZF\_PHD\_2; 3.  
 KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;  
 KW Nuclear protein; Developmental protein; Activator;  
 KW Alternative splicing.  
 FT ZN\_FING 1266 1347 PHD-TYPE 1.  
 FT ZN\_FING 1348 1393 PHD-TYPE 2.  
 FT ZN\_FING 1421 1482 PHD-TYPE 3. (ATYPICAL).  
 FT ZN\_FING 1734 1793 PHD-TYPE 4 (ATYPICAL).  
 FT ZN\_FING 1794 1844 PHD-TYPE 5 (ATYPICAL).  
 FT DOMAIN 3599 3708 SET.  
 FT DOMAIN 512 516 POLY-SER.  
 FT DOMAIN 565 570 POLY-ASP.  
 FT DOMAIN 661 664 POLY-SER.  
 FT DOMAIN 905 910 POLY-SER.  
 FT DOMAIN 1576 1582 POLY-SER.  
 FT DOMAIN 2298 3027 GLN-RICH.  
 FT DOMAIN 3032 3040 POLY-SER.  
 FT DOMAIN 3181 3184 POLY-GLN.  
 FT DOMAIN 3220 3225 POLY-GLU.  
 FT VARSPLIC 1 368 MISSING (IN SHORT ISOFORM).  
 FT CONFLICT 2025 2025 P -> PWLSPLEFLGISTHGILLWLLGCVVRLKGG  
 (IN REF. 1).  
 FT CONFLICT 2341 2341 R -> S (IN REF. 1).  
 FT CONFLICT 2392 2392 G -> S (IN REF. 1).  
 SQ SEQUENCE 3726 AA; 400575 MW; D275650763D1CF5 CRC64;  
 Query Match 2.8%; Score 133; DB 1; Length 3726;  
 Best Local Similarity 18.3%; Pred. No. 63;  
 Matches 156; Conservative 122; Mismatches 281; Indels 292; Gaps 43;  
 QY 13 MPVALAAYPLMYS-----QALAQNNPANIINHVAHDATNAQAKGNPVLTPEDQI 66  
 DB 2997 MPITLLEAPVQSQFMEVPEQALEQ--ELANRVQHFSTSSSS--SSNCSLPTNVVNPQ 3054  
 QY 67 QARLNAGLNAPQSQQLDVNFDQSPISR----IGQSPRLQDMSVIEETPLSLEE 122  
 DB 3055 QABSTTSSSTRPTNRLPMQQRQEPAPLSNECPVSSPTPKPEQPTIHQMTSASVS 3114  
 QY 123 LFAQESTEMGINPDYIPREYQGEQPNSEV--VVP-----PTLEPEKGLIKRLVARLND 175  
 DB 3115 CYAQKST---LPS-----PYVEALIKVSSVLESITPDVTMAILEQP-----VQSITYE 3162  
 QY 176 GVNVKPLRAKFKYQSSQSGETSAIGSHQTEPYANIKALEDTQGSAMDNGSIPRLR 235  
 DB 3163 GL-----YKNSPGSS-----KTEQLLLQQQREQLNQ--LVNNG----- 3196  
 QY 236 QTAIVARAQVYIDLSIIRNSIGEVYIHDGEVYIDRAVEVRGEGADKAF--- 292  
 DB 3197 -----YLDKHTTQVEPMDIV-----YREEDLEEEDDDDFSLK 3232  
 QY 293 --TTVADEVPLTIGDVFNHGKYEFTKKMLIENASAEHGYFDGRWLDRSVVILPNTADV 350  
 DB 3233 MKTSACND-----HEMSDESEPAVKDKIK--ILDNLTDDCADSIATITMEVD 3280  
 QY 351 LIYDTGTQVRFDEVVFETIDPKTNLTTPDKLPVKRELLQLLTVNNGEAYNLQAVRAL 410  
 DB 3281 A--SAGYQWVEDVL-----ATTAQSAF-----TEFEGLAETRAVEA- 3317  
 QY 411 SNDLIATRYRN-WNNTIVEPPEQIOND-QVSEQSSSTTEPAQVDESLIEVETVE 468  
 DB 3318 ----AATYINEMADAVL--DLKQIQNGVELLRKRREKQRTVSQBOESKAIVPTAA 3370  
 QY 469 LTDCILMDISPIEFSASNLIDOKLNLVAAKARHLYDMPDRVLAINHDDGVN-----RS 522  
 DB 3371 APEPQPIQEPKMTGPHL-----LYE-----IOSEDDFTTYSSTITE 3408  
 QY 523 ILGRISDAVASAVARAILPDESENEVIDLP-----RTALANKRTADVYQ----- 567  
 DB 3409 IMEKVFEAVOVARRA-----HGLTFLPGLADMGIGIMIGKTALKYLILQPLGVE 3461  
 QY 568 --SKKVPVLY-----VFVADKPRDQIG-----LNGSDPTGTLVLT 601  
 DB 3462 KCSKYPPIKHKRNGNVSTANGAHGNLGGSSASALSVSGDGHLLDYGSDQ-----D 3516

```

OY 602 KREHNL-----NRDYGAGELRLSEDKKGYLATKPLSHPLNDOLRATLGYYOQ 652
DB 3517 ELEENAYDCARCEPYSNRESEYDMFSLASRHRKQPIQVYV-----QPSDNELVPRRG----- 3568
OY 653 EYFGSTNCFDLSTRLLEHRSRIIONGGMNRTYSLKRYLRLKLTQADPEWMDLPYDF 712
DB 3569 -----TGSNLPMAKRYRLK-----ETRYKDYGVGF 3593
OY 713 ---VNGK-----PSEFALL--AGVAVHKTVDNLVPMRGYRORY--SLEVG----- 752
DB 3594 RSHINGRGICXKDIKDEAGMAYIEYAGELIRSLTLDK-----RERYDSRGIGCYMFK 3645
OY 753 -SSGLVSDAMN 762
DB 3646 IDDNLVVDATM 3656

RESULT 21
CLPB_HAETN STANDARD: PRT: 856 AA.
AC P44403;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE CLPB Protein.
GN CLPB OR H10859.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shetty R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR
CC ATP-DEPENDENT PROTEASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: U32767; AAC2518.1; -
DR TIGR: H10859; -
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003959; AAA_ATPase_cent.
DR InterPro: IPR001270; Chaprinin_c1pA/B.
DR InterPro: IPR004176; C1P-N.
DR Pfam: PF00004; AAA_2.
DR Pfam: PF02861; C1P-N; 2.
DR PRINTS: PR00300; CLPROTASEA.
DR PRODOM: PD000739; GSPIL_E.1.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS00870; CLPAB_1; 1.
DR PROSITE: PS00871; CLPAB_2; 1.
KW Chapterone: ATP-binding; Repeat: Complete proteome.

```

```

FT DOMAIN 161 409 I.
FT DOMAIN 531 722 II.
FT NP_BIND 206 213 ATP (POTENTIAL).
FT NP_BIND 605 612 ATP (POTENTIAL).
SQ SEQUENCE 856 AA; 95837 MW; B8958ED9BD03EA3B CRC64;

Query Match
Best local similarity 19.3%; Pred. No. 7.1; Indels 147; Gaps 25;
Matches 109; Conservative 96; Mismatches 212;

OY 28 ALAQONNPANIIINHVDPAHDTAINQAKAGNPPVLLTPEQIARLNAAGLNKPPQALDVV 87
DB 115 ALEERGTISDILKKCGAKKEQISOA-----I0HIRGGQNVNDQNAEERQALEKY 164
OY 88 NPDQSPISRGQSPPLGLDMVIEETTPLSLEELFAQESTMGINDPNIPRYGGEOP 147
DB 165 T1-DLTARAESGKLDPVIGRD-----EEI-----RRATQVLQRRRK 199
OY 148 NSEVVVPTLEP--EKGGLIKRLYARLFNDGVNKKVPR-LKAKFYOSSQGETSATGSSHQ 204
DB 200 NNPPVLIG---EFGVGKTAIVEGLAQRIVN---GEVEGLKNNKRVLSLDMG--ALIGAKY 251
OY 205 KTEPPANIKALEDITQESAMDINGSIPLRQTAIVAAVAGYDIDLSIRNSIGEVY 264
DB 252 RGEFERLKAVALNELSKEGRVIL-FIDEI-HTMVGAKTDGAMDAG-NLKRSLARGE- 307
OY 265 ITHDGEPIYIDYRAVEVGEADGKAF-----TTVADEVPLIG-----DVFHGKY 312
DB 308 -IHCVGATTLDERGY-IEKDAALERFQKVFVDEPSEVDITALEIKERYEIHIIHVDI 365
OY 313 ETKKNLIENASAEHGFYDGRWLDRSVYDILPNTADVSLIYDTGTQYRDEVVFTIDPK 372
DB 366 -TDPAIVAAATLSHRYISRQ-----LPDKALD-----IDDA 397
OY 373 TNOLTTPDPKLPYKELLEQLLTVNGEAYNLQAVFALSNDLIARFYFMKVTEIPEPR 432
DB 398 ASSIMEIDSKPEPLDRERRITQLEQALQK-----EEDBASRRLEMLKELEAKER 453
OY 433 EGIQNDQY-----SFEQSSSRTEPAQVDES-----TLEPYI 464
DB 454 EYAELEEVWKSKAATLSSGQHIKQELDTAKTELEQARRRGDIKAKSELOYGRIPDLKOL 513
OY 465 ETVELTDGILM-----DISPIEFSASNLIDKLLVAAKARHLYMDPPDRLAIN 514
DB 514 EQAERSSEKEMTLKRYVTDIEIAVLKATGIPYSKM-----MGEKEKLR 561
OY 515 HDDGYNRSILGRISDAVSAVARAI 538
DB 562 MEDELKRRVIGQ-ERAVDAVANAI 584

RESULT 22
YD86_SCHPO STANDARD: PRT: 1957 AA.
AC Q10411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein Clf3.06c in chromosome I.
GN SPAC13.06c.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sounos J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,

```

RA Gentes S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,  
 RA Goffeau A., Cadieu E., Driano S., Gloux S., Lelaure V., Mottier S.,  
 RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revelante J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
 RT The genome sequence of *Schizosaccharomyces pombe*.  
 RL Nature 415:871-880(2002).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Z70690; CAA94624.1; -  
 KM Hypothetical protein.  
 SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 2.8%; Score 132.5; DB 1; Length 1957;  
 Best Local Similarity 18.1%; Pred. No. 25;  
 Matches 168; Conservative 149; Mismatches 322; Indels 289; Gaps 39;

OY 17 LAAYPLPLMTSQLAQQNNPANIINHVA-----HDTAINQAKAGNPVLLTPDIOIA 68  
 DB 764 LTAANVETLTQDSKAMQSTSLVNSQISNLXHELRDQVNNQSON-----TLSES 818  
 OY 69 RLNAAGLNAKPPSOAL-DVY-----NPDQSPISRIQSGSPPLGLDMSVIEETPLSLE 121  
 DB 819 KLKTDCENTLQONMFLIDVQKIMKHVNOESKVSLEKENVCKSLDKNL----- 869  
 OY 122 ELFAQESTEMGINPDYI-----PEYQGEQNSEVYVPTLEPEKGL---I 165  
 DB 870 -----RSSLVNAISDNDQILTQLAEISKVNSLEQESAQLNSGL---KSLAEAKQLHTEN 922  
 OY 166 KRLVARELNDGVNKP-RLKAKFYOSOGSETSAIGSSHQKTEPYANIFALEDITQESA 224  
 DB 923 EELHTRL-----DKLTGKAKIESSSDGKLT-----ARQELSNLKE--ENMSQSA 970  
 OY 225 MDLNSIRLRQTLVAARAVGYIDISITNSIGEDVY-----ITHDL---G 270  
 DB 971 ITSVKS--KLDELTSSKSK---LEADIEHLKNKVSVEVERNNALLASNERMLDKNNG 1024  
 OY 271 EPPYIDYRAVEVBGAGD--KATTVADVP--LLIGVFHNGKETKKNLIENASAEH 326  
 DB 1025 ENIASLQTEIEKKRAENDLOSLSVSEYENLLIS-----QTNKSLEDKTN--- 1074  
 OY 327 GYFDGRMLDRSDVYLIPN-----TADVSLYPTGTGYRDEYVFFITDPKTNL 376  
 DB 1075 ---QKYTEKNVQKLLDKDQANVELELTISKYKGLGEENAOIKDELLARKKSKKHDL 1131  
 OY 377 TTD-PDKLPVRELLEQLITVNGEAYNLQAVRALSNDLIARRYNMVTEIVFEREQI 435  
 DB 1132 CANFYDDLKEKSDALEQLTNE-----KNELIVLEQSSNSNNEALVVERSDL 1177  
 OY 436 QNDQVSEFSSSSRRPEAQNVESTLEPIYETVELTDGLIMDISPIEFSSNLIDQK---- 491

DB 1178 ANRLSDMKKSLSDSDNIVSIRSDLVRYNDEL---DTLKKDKDSLSTQYSEVCODRDDLL 1234  
 OY 492 -----LNLVAKARHL-----YMPDRVLAIN----- 514  
 DB 1235 DSLKGESEFNKAVSVIRLCTKSEIDVPVSEIILDDNFVFNAGNFSLSRLTYVLSENYL 1294  
 OY 515 -----HDDGVNSILGRISDAVSAVARA 537  
 DB 1295 DAFNQVFNKMEIDNRLTTTDAEFTKVVADEKLQHEHDMWLQR--GDLEKALKDSEKN 1352  
 OY 538 ILPDESE-NEVIDLPETALANRTPADYQSKVPLVYVASRPPRQDQIGLWGSPTG 596  
 DB 1353 FLRKAEMTENIHSLEGGKETKEIAEL-----SSRLDENQIAT----- 1392  
 OY 597 TRLVTKREHNLINDGYQAGAEIRLSD-----KKGVLVYATKP 635  
 DB 1393 NKLKNQDHL--LND-----EIRLKEDVLKEKESLIISLESLSNOROKESSILDKNE 1443  
 OY 636 LSHPLNDQLRATLGYQOEVEGHSTN---GFDLSTRTLE-----HEISRIQNGWN 684  
 DB 1444 LEHMLDTSRKNSSIMKISINSISLDDKSPELASAVKGLALOKLSESLSMEN----- 1499  
 OY 685 RITSIRLRD--KLKTQAPPETWODLPVDFVNGKPSOE-----ALLAGVAHKTVADN 735  
 DB 1500 -----IKSQLQAEKKEIQVDESTIOELDHEITASKNNYEGKLNKDRIIRLSENIQLNN 1555  
 OY 736 LVNPMRGYRQRYSEVSSGLVSDANMA 763  
 DB 1556 LLAEEKSAVRSLSTEKESEILQFNRI 1583

# RESULT 23

HS75\_KLUMA STANDARD; PRT; 612 AA.

AC P41770;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Heat shock protein SSB.

GN SSB.  
 OS Kluveromyces marxianus (Yeast) (Kluveromyces fragilis).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluveromyces.  
 NC NCB1\_TaxID=4911;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12424;  
 RA Iborra F.

RL Submitted (MAY-1991) to the EMBL/Genbank/DDJ databases.  
 CC -!- FUNCTION: MAY AID IN THE PASSAGE OF THE NASCENT POLYPEPTIDE CHAIN  
 CC THROUGH THE RIBOSOME CHANNEL INTO THE CYTOSOL. SUCH AN INTERACTION  
 CC COULD BE CRUCIAL FOR CONTINUOUS TRANSPORT OF THE POLYPEPTIDE.  
 CC COULD SERVE TO PREVENT THE NASCENT POLYPEPTIDE FROM INTERFERING  
 CC WITH TRANSLATION BY CLOGGING THE RIBOSOME CHANNEL (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH TRANSLATING  
 CC RIBOSOMES. MAY BIND DIRECTLY TO THE NASCENT POLYPEPTIDE.  
 CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X59963; CAA42589.1; -  
 DR HSP; P08107; IHUO.  
 DR InterPro: IPR001023; HSP70.  
 DR Pfam: PF00012; HSP70; 1.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR ProDom; PD000089; HSP70; 1.  
 DR ProSITE; PS00297; HSP70\_1; 1.

DR PROSITE; PS00329; HSP70\_2; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 KM Heat shock; ATP-binding; Multigene family; Protein biosynthesis;  
 ACetylation.  
 FT INIT\_MET 0 0 BY SIMILARITY  
 FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY)  
 SQ SEQUENCE 612 AA: 65955 MW: 571881571A78410 CRC64:

Query Match  
 Best local similarity 19.9%; Pred. No. 4.5;  
 Matches 134; Conservative 94; Mismatches 267; Indels 180; Gaps 33;

2.8%: Score 132; DB 1; Length 612;

QY 81 SQALADV-----NFDDQ-----PLSRIGESPPLGDMVIEETPLSLELFAEQ 127  
 DB 66 NTVFADAKRLIGRFDEESVSDMKTPFKYIDSGAPL-IEVELGETKITSPOEISSMV 124

QY 128 STEMG-----INPDYIPEYQEQPNS-----EVVPPTL----- 157  
 DB 125 LTKMKIEAEKIGKKEKAVVTPAYFNDAGROATKDAAGTAGLNLRLINEPTAAATAY 184

QY 158 -----EPEKPGILKRLYARLFN-----DGVNKPRLKAKFYQSSQSETSAIGSSH 203  
 DB 185 GVGAGNSEKEKHVILFDLGGTFVSLHLTAGVYTVKSTSGNTHLGQDDPTNL- 242

QY 204 OKTEPYANIKAALEDITQESAMDNGSIPRLRQALVAARAVGYDIDLSIRNSIGEVD 263  
 DB 243 FKTE-----FKKTKGADISGABARLRRLTAERA-----KRTLSVAQTVEVD 287

QY 264 VIHDLGEPYIDYRAVEVREGADDKA--FTYADEVPLLIGVFHHGKETKKNLIEN 321  
 DB 288 SLFD--GE--DEAATTRARFEDINAALEFKSTLEPEVOYLKDA-----KISKQIDE 335

QY 322 ASAEHG-----FEDGRMLDRSVVILPDNTADVSLYDGTQYRFDEVEFT 368  
 DB 336 VLVGSGSTRIPKVKQLSDPFQKQLEKSN-----PDAVAVGAAYQAT----- 381

QY 369 IDPKTNLTTDPDKLPVKRELLEQLLTVMN-GEAVNLQAVRALSNDLITRYFMNV--N 424  
 DB 382 ---LTGQSTDETKDLLLVAPLPLSGVMAGDFGVVPPRNTTVPRTIKRRTTVAHQ 438

QY 425 TEIYFPERQIQNOVSFEQSS-----SSRTEPAQVDESTLEPIE-----TYE 468  
 DB 439 TYVTFP---VYQGERVCKEHTLLGEPDLKGVPPMAGEPVLEAIFEDANGILKTVAVE 495

QY 469 LTDGLMDISPIEFSASNLIDQKLNLAARARHLYDMPDDEVLAINHDDGVNRSILGRIS 528  
 DB 496 KSTGKSANIT-ISMALIGRLSEELIEQVNVQAEF--KAADFAFAKKH-----ARQRL 546

QY 529 DAVSAVARAILPDSENEVIDLPERTALANKRP-ADVYOSKAVPLVYVASDKPRDQI 587  
 DB 547 SYISSVOQYITDPLVSAKI-----KRNKAKAVEALADAFSTLQIE--DASADDLRKA 599

QY 588 GLGWGSDTGRLYTK 602  
 DB 600 GL-----KRAVTK 607

RESULT 24  
 RASO\_THEAC  
 ID RASO\_THEAC STANDARD: PRT: 896 AA.  
 AC 09HLR8:  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA double-strand break repair rad50 ATPase.  
 GN RAD50 OR TA0157.  
 OS Thermoplasma acidophilum.  
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;  
 OC Thermoplasmataceae; Thermoplasma.

OX NCBI\_TaxID=2303;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=DSM 1728;  
 RX MEDLINE=20479972; PubMed=11029001;  
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;  
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma  
 acidophilum";  
 RL Nature 407:508-513(2000).  
 CC -i- FUNCTION: Involved in DNA double-strand break repair (DSBR). The  
 CC rad50/mre11 complex possesses single-strand endonuclease activity  
 CC and ATP-dependent double-strand-specific exonuclease activity.  
 CC Rad50 provides an ATP-dependent control of mre11 by unwinding  
 CC and/or repositioning DNA ends into the mre11 active site (by  
 CC similarity).  
 CC -i- SUBUNIT: Forms a complex with mre11 (By similarity).  
 CC -i- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AL45063; CAC11304.1; -  
 CC InterPro; IPR003439; ABC\_Transport.  
 CC InterPro; IPR003395; SMC\_N.  
 CC Pfam; PF02463; SMC\_N; 1.  
 CC DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.  
 FT NP\_BIND 30 37 ATP (By SIMILARITY).  
 FT DOMAIN 151 744 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 896 AA: 103435 MW: 2C584D700CCD2CB2 CRC64;

Query Match  
 Best local similarity 19.2%; Pred. No. 8.7;  
 Matches 157; Conservative 122; Mismatches 297; Indels 243; Gaps 36;

2.8%: Score 131.5; DB 1; Length 896;

QY 79 POSQALDVNFDDQSPISRIEQSPPLGDMVIEETPLSLELFAOESTEMGINPDY 138  
 DB 93 PESNAMIW---DGSALQSQVANDY-IEKNILTKSKOVFLNSVSKG-EM---DDL 143

QY 139 IPEYQGEQNPSEVVPPLLEPEK-----PGLIKRLYARLFN-----DGVNKPRLKAK 186  
 DB 144 I---SGDARRKKLLDELTLEKLEETVYDKVDLSIQAGISNLDYLISENERDDLR 200

QY 187 FYQ-----SSQSGETSAIGSS--HQTEPYANIKAA-----LEDITQESAMDNGSIPR 233  
 DB 201 RYQDDVAELSKQIQOEAIEEDLKRKEASAEVYAAVSKELIMDATLKNMSSLDEANR 260

QY 234 LRQTL-----VAARAVGYDIDLSIRNSIGEVDVIHDLGEPYIDYR----- 278  
 DB 261 YEERIRKIDKLOEISGSTERNEYETSSKYVASRERIGYWDKQ--IIDYRKMLKNID 318

QY 279 -----AVEVREGADDKAFTTVADVPPLLIDV-FHHGKYETKKNLIENASAE 325  
 DB 319 GQVQSYEDNMKKAELQADHQYELMQRMDIEKLEDDLRKYESKYYSLINEIQKKK 378

QY 326 HGYPDGRWLDRSVVILP-----DNTADVSLYDGTQYRDEVVEFTIDPKTNLTTDP 380  
 DB 379 REEYRKQKQDGLDEISRLGRAFAVASLVAIYE---EIRRD-----IDEINPDL 425

QY 381 DKLPYKRELFO-----LLTVNMGAVYMLQAVRALSNDL-----IATRY---FNMVNT 425  
 DB 426 GNLYVKIGALKQKEEIRRNMMLEGNHKCPV--CGTLDGEGSRIRIEHSEDINRNE 483

QY 426 EIVPPER-----QTONOVSEFGSSS 447  
 DB 484 EIDHLERASAIDEKKROLISMESYLAKGKIREYETVROKKLEAQITTDENSLSITAY 543

QY 448 SRTEPAQVDESTLEPIETV--ELTD-----GILMDISPIEFSASNLIDQKLNLAARAR 500

```

Db      544 KHTKTEQJDEEERSHHLEDLKQKTTDMNNAVAVNSIGDIE-----ALRKOKEVSKKL 598
QY      501 HLYDMPDRDLVAINHND-----DGVNKSILGRISDAVSAVARAI-LPDESENEVIDLPERTA 555
Db      599 D---AEDRTEHEIESEFPDINSYTPSYIGKTEDEVRILLEPQIKLAEDLKQRETLRKKVK 654
QY      556 -LANKTTPADVYQSKKVPVLYFVAFSDKPR-----DGOJGLGMSGSTGRILVTKFEHNLINR 610
Db      655 DLRSSACMDIEIQRRKNELSVKASESEFRLRKYVGEQIOATLSSLSGR--SKVE---TL 708
QY      611 DGYQGAELRESEDKKGVKLYATPRPLSHPLNDOLRATLGVQOEVEGSHSTNGFDLSTRTLE 670
Db      709 RSHVSEIQRIQSDRERDIE--RMKKIEAINDVKRI-----REAFGR----- 748
QY      671 HEISRSIIQNGGMNRTYSLRYRLDKLTKQAPEPTEMODLPYDEVGKRSQCEALLAGVAHVH 730
Db      749 -----NGVPRAM-----IRQ 757
QY      731 TVADNLVPMRGYRQORYSLEVGSSGLVSDANMAIARAGI 766
Db      758 SVSDDLAKTAKTDYLSSEFDLDDISVQODFNWTVYVRGCV 796

```

RESULT	25
RPOB_UREPA	
ID	RPOB_UREPA
AC	Q9PQV6;
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	DNA-directed RNA polymerase beta chain (EC 2.7.7.6) (Transcriptase beta chain) [RNA polymerase beta subunit].
GN	RPOB OR U0187.
OS	Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX	NCBI_TaxID=134821;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Serovar 3;
RX	MEDLINE=20500219; PubMed=11048724;
RA	Glass J.I., Lefkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,
RA	Cassell G.H.;
RT	"The complete sequence of the mucosal pathogen Ureaplasma urealyticum".;
RL	Nature 407:757-762(2000).
CC	-I- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS SUBSTRATES (BY SIMILARITY).
CC	-I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).
CC	-I- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1 BETA' CHAIN (BY SIMILARITY).
CC	-I- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; AE002118; AAF30594.1; -
DR	HSSP; Q9KMU7; IHQM.
DR	InterPro; IPR001572; RNA_POL_B.
DR	Pfam; PF00562; RNA_POL_B.1.
DR	PROSITE; PS01166; RNA_POL_BETA.1.
KW	Transferase; Transcription; DNA-directed RNA polymerase;
KM	Complete proteome.
QC	SEQUENCE 1434 AA; 162022 MW; 2841BFA50C19BA7A CRC64;

Query Match	Similarity	2.8%	Score 131	DB 1	Length 1434
Best Local	Similarity	19.2%	Pred. No. 19		
Matches 158	Conservative 132	Mismatches 281	Indels 254	Gaps 43	
OY	56	NPPLVLTPEQIOARLNAAALLNAAK-POSQALDVVNFDDQSPISRIQSGPGLDMSVE	114		
Db	641	NPLAELTNNRRIRISAMOPGSIREDRPLRIDVY---HHSHSRFLCTETPEEGNIGLI--	694		
OY	115	TPPLSLEELFAOES-----TEMGINPNDYIPREXOGEOPNSEVNVPP--LEPEK	161		
Db	695	---MSIASLAKVDENGFIYAPRYVEDGVKEDY--KYTLAHEDBDYIIAESSYQDENK	749		
OY	162	PGLIKRLAR-----LFN---DGVNKKPKRLAKAFYQS-----SQSETSAIGSSHOK	205		
Db	750	RILDDQVARYARGSTGLFSPNEVDLIDYPRKVVNSIAASAIFETIENDDCARALMSNMOR	809		
OY	206	-----TEPAANKALED--ITQESAMDLNLSIPRLRGALVAAR--ANGYDIDLSTIR	256		
Db	810	QATPLIKPAPRIYVGCTETEKRIAHDSG-----AYAAKDDGVVEFVDSOKIIR	857		
OY	257	NSIGEVLDIINDLGEPPYIDYRAVEREGADKAFITVADEVPLL-----IGD	305		
Db	858	NNDKRD-----DYKLITYRKSNQD-----FCNNQIPLVKGVQRVHKSETIGD	900		
OY	306	---VFHHGKYEYTKNLLENASAEHG-----FDGRMLDRSDVLL-----	342		
Db	901	GRAMONGELATLRNLVGYTWRGYNFEDAILISRLYDQDVFTSIHIDEHTIQCKMTKN	960		
OY	343	---PDMTADVLSLYDT-----GTOYRFDEVEFFITDPTNQLTTPDRL---	383		
Db	961	GDEETTRDMPVNSDTAKRFLDNOGLYLVGAENHESDVLGKTTPRGNVETAPEDRLQTI	1020		
OY	384	--PVARELLEOLLTVNMGAEAYNLQAVR-----ALSNDLIATRYFNMYNTEIPEPE	431		
Db	1021	FEGDSKTYKQDSSSLAYKHGQEGIVAAVNRKIKSSDENGSELPRDYI-----EIKYIV--Q	1073		
OY	432	RQOIQ--NQVSEQSSSSTPEPAQVDESTLEPVITVELTGDILMDIS-----PIEFA	484		
Db	1074	KRIIOGVGRKMARGRHNKGIV-----SKVVPIDQMPFLKDTPTDIPMLNPLGVPSSRMNI	1126		
OY	485	SNLIDDKNLVNAK--ARHLIDMPDRVLAIHND-----GVNRSILGRISDAVAVARA	537		
Db	1127	GOELHLHGTYAAAEIGKKQLOIDAQ---LGTEKYISLFGINELIARKLVENISNLTKH	1183		
OY	538	ILPDESENEVDILPERTALANKRTPADYQSKKVPFLYFVASDKPRDQOIGLGWGSPTGT	597		
Db	1184	KQAKQAKD--IDLIDVYTIILK-----ELGLSTY--DDIGI	1213		
OY	598	RLVYKF-----EHNLI-----NRDGYQAGAEILRLSEDKG--VKLYATKPLSH	638		
Db	1214	KISTPVEFGANHHDDIVSIIMEANIDIEENKKGVOLYDQRTGPRPGLISVGLTYMLKIDH	1273		
OY	639	PLINDLRA-TIG-----YQAEVFGHSTNGFPLSTTLEHLSRSRLIQNGGN--RTYSIRY	691		
Db	1274	MVDDIHSRVSQVSKITQOPLGGSQNG---GQFFGEMEV-----WALEAYGAAY	1321		
OY	692	RLDKLKTQAPPEYMODLPVDFVNGKPSOEALLAGVAVKTYADNL	736		
Db	1322	NLELIT-----IKSDVQGR--NQATYNAIIKGHDVYADGM	1355		
RESULT 26					
SCA4_RICE					
ID	SCA4_RICE	STANDARD:	PRT:	981 AA.	
AC	OGA37:				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)				
DE	(Protein PS 120) (fragment).				
GN	SCA4 OR D.				
OS	Rickettsia felis (Rickettsia azadi).				
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;				

CC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=42862;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sekegova Z., Roux V., Raoult D.;  
 RT "Phylogenetic analysis of Rickettsia spp. by comparing the sequence of  
 RL gene D coding for an intracytoplasmic protein."  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AF196973; AAK31304.1; -  
 KW Antigen.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 981 AA: 107514 MW: 7F18F421E2C262E1 CRC64:  
 Query Match 2.88; Score 130; DB 1; Length 981;  
 Best Local Similarity 18.4%; Pred. No. 12;  
 Matches 166; Conservative 136; Mismatches 348; Indels 252; Gaps 39;  
 OY 62 TPEDIOARLNAGLNAKQSOALDVNFDDOS-PISRIGESPP-GLDSVI-----EE 114  
 DB 23 TPQLEEA---DDGFSFTPASSSTQSPISLSSGSSSDQSDPTTKAVRETIIOPOKDE 79  
 OY 115 TTPSLSELEFAQESTEMCINPDYIPEYQGEQNSEVVVPTLEPEKPKILRIYA--- 170  
 DB 80 IAEIOLKDLALADRLD-----AEQKRKEI-----EEEKDKTISAFGNPAN 121  
 OY 171 RLFDNGVVKVRLAKFAFQSSOSGFTSAIGSSHOKTEPYANIKALEITQESANDLNGS 230  
 DB 122 REFDKALENELKLLKLEISIEAGKNVL-STYSANGYQGFQVQENQISASDLNAT 180  
 OY 231 IPR-----LROTAL-----VAARAVGYVIDLSIIRNSIGEVDYIIDLGEPIYD 276  
 DB 181 VVRNDAGDELCTLNETTYKTKPFYAKQDG-----TQVQINSYRIDP-----PIKLD 228  
 OY 277 -----YRAVEVGEG--ADDKA--FTYVADEVPLLIGVFHHGKETKKNLIENASA 324  
 DB 229 KADGSMLSMVALKADGKPKSKDAVFTAHYEEGP-----NGKPOLK-----EISSP 276  
 OY 325 EHGFFDGRMLDRSYDVILPONTADVSLYDTGTQYRFDEVVFFITDPKTNOLTTDPDKLP 384  
 DB 277 KPLKFAG-----DGPDAVATIEHGEIY-----TLA 302  
 OY 385 VKRELEQL---TVNNGEAVNLQAVRALNSDLIATRYFNWNTETVEPEREOIONDOVS 441  
 DB 303 VTRGKYKEMREVELNGOSVDLS--QTIADLTQVGRSOETPOPIITTPQDELKS---S 357  
 OY 442 FEQSSSRTPEAQVDESLPEVIEFTVETDGIIMDISPIEESASLIDOKLNLVAAKARH 501  
 DB 358 IETPTTVOPIITPANOPLOPETSQMPQOVNPNPLVATGSLSTMODLLNYVA--- 413  
 OY 502 LYDMPDDRVLAINHDDGVNRSILGRISDAVASAVARAILPDESEN-----EVIDLPERTA 555  
 DB 414 -----CLTRKESGNKOI-----DLINEAATAILNNKESDIAEKQANIIALTENTV 458  
 OY 556 LANRKP-----ADVYQSKKVPPLYEVA-----SDKPRDGOIGLGW 591  
 DB 459 NNNDLTPTDVKAGVNAVLETIKNDQNPDELSKSMLEATVAIALNSEQEKQKQOMLEK 518  
 OY 592 GSDVGTGLVYKFEHNILNR-DGY-QAGAEALRSDDKKVKKYALKRPLSLNDQLRALGL 649  
 DB 519 AVVDGLSL--KDDASRYVAIDGITDAVAKSNLSTEDKOTMLAV----- 560  
 OY 650 YQGEVFGHSTNGFDLSTRTLEHEISRSIIIONGNRTYS-----LRYRLDKLTQAPPE 703

DB 561 -----CDKVNASSELNAE-KOKLLGSLVKKGVAEQVLSPEQOOLMOQNLDKI-TAEQTK 612  
 OY 704 TWQDLPVDFVNGKPSQDELLAGVANHKTVADNLVNPMPMGYRGQRYSLVEGSSGLSDAMA 763  
 DB 613 NAQITVEYGLIANPENTIAKTEALQNTVTKVLSDFPIAEIKGETLE-----SITKV 664  
 OY 764 IARAGISG-----YSPGDNAVGSNRAHOMTGIGIAGYIWSDNENHVPYRLRFPAGDQ 817  
 DB 665 VAESPLNGQDKRADIVKNGE-ALASHKTMAPTEKIST-----IE 702  
 OY 818 SIRGAHDSLSISDKGYLTGGQVLAAGTAETNPFKMDRLAVFGDIGMAYDKFTMDT 877  
 DB 703 SVEKVAESTIDLEBKRLMTGLVGEIGKRNPEITSEKTKAV-----SRGIDKST 754  
 OY 878 KI 879  
 DB 755 AI 756  
 RESULT 27  
 PGCV\_RAT STANDARD: PRT: 2738 AA.  
 ID PGCV\_RAT  
 AC QPERBA: 008592; 088564; QPR1K4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Versican core protein precursor (large fibroblast proteoglycan)  
 DE (Chondronatein sulfate proteoglycan core protein 2) (PG-M) (Glia  
 DE hyaluronate-binding protein) (GHAP) (Fragments).  
 GN CP6G2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0), SEQUENCE FROM N.A.  
 RP (ISOFORM V3), AND SEQUENCE OF 2657-2738 FROM N.A. (ISOFORM VINT).  
 RC STRAIN=Wistar Kyoto;  
 RX MEDLINE=99327053; PubMed=10397680;  
 RA Lentre J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,  
 RA Wright T.N.;  
 RT "Versican/PG-M isoforms in vascular smooth muscle cells";  
 RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).  
 RN [2]  
 RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0).  
 RC STRAIN=Wistar Kyoto;  
 RX MEDLINE=98308094; PubMed=9642104;  
 RA Miley P., Maurel P., Chiba A., Mevissen M., Popp S., Yamaguchi Y.,  
 RA Margolis R.K., Margolis R.U.;  
 RT "Differential regulation of expression of hyaluronan-binding  
 RT proteoglycans in developing brain: aggrecan, versican, neurocan, and  
 RT brevican";  
 RL Biochem. Biophys. Res. Commun. 247:207-212(1998).  
 RN [3]  
 RP SEQUENCE OF 2421-2463 FROM N.A. (ISOFORM V0).  
 RC TISSUE=Kidney;  
 RX MEDLINE=98094159; PubMed=9434070;  
 RA Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;  
 RT "Proteoglycan expression in the normal rat kidney";  
 RL Nephron 77:461-470(1997).  
 RN [4]  
 RP SEQUENCE OF 2535-2738 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Lung;  
 RA Blomberg L.A., Chan W.-Y., Clerch L., Massaro D.;  
 RT "Molecular cloning and characterization of two developmentally  
 RT regulated genes in rat lung";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: May play a role in intercellular signaling and in  
 CC connecting cells with the extracellular matrix. May take part in  
 CC the regulation of cell motility, growth and differentiation. Binds  
 CC hyaluron acid.  
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.

FT	DISULFID	2440	2455	BY SIMILARITY.
FT <th>DISULFID</th> <td>2457</td> <td>2466</td> <td>BY SIMILARITY.</td>	DISULFID	2457	2466	BY SIMILARITY.
FT <th>DISULFID</th> <td>2473</td> <td>2484</td> <td>BY SIMILARITY.</td>	DISULFID	2473	2484	BY SIMILARITY.
FT <th>DISULFID</th> <td>2478</td> <td>2493</td> <td>BY SIMILARITY.</td>	DISULFID	2478	2493	BY SIMILARITY.
FT <th>DISULFID</th> <td>2495</td> <td>2504</td> <td>BY SIMILARITY.</td>	DISULFID	2495	2504	BY SIMILARITY.
FT <th>DISULFID</th> <td>2511</td> <td>2522</td> <td>BY SIMILARITY.</td>	DISULFID	2511	2522	BY SIMILARITY.
FT <th>DISULFID</th> <td>2539</td> <td>2631</td> <td>BY SIMILARITY.</td>	DISULFID	2539	2631	BY SIMILARITY.
FT <th>DISULFID</th> <td>2607</td> <td>2623</td> <td>BY SIMILARITY.</td>	DISULFID	2607	2623	BY SIMILARITY.
FT <th>DISULFID</th> <td>2638</td> <td>2681.</td> <td>BY SIMILARITY.</td>	DISULFID	2638	2681.	BY SIMILARITY.
FT <th>DISULFID</th> <td>2657</td> <td>2694.</td> <td>BY SIMILARITY.</td>	DISULFID	2657	2694.	BY SIMILARITY.
FT <th>CARBOHYD</th> <td>57</td> <td>57</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	57	57	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>330</td> <td>330</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	330	330	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>692</td> <td>692</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	692	692	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>758</td> <td>758</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	758	758	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>802</td> <td>802</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	802	802	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>805</td> <td>805</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	805	805	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>1257</td> <td>1357</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	1257	1357	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>1435</td> <td>1435</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	1435	1435	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>1633</td> <td>1633</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	1633	1633	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>1660</td> <td>1660</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	1660	1660	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>1684</td> <td>1684</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	1684	1684	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>1738</td> <td>1738</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	1738	1738	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>1848</td> <td>1848</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	1848	1848	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>2004</td> <td>2004</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	2004	2004	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>2409</td> <td>2409</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	2409	2409	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>2711</td> <td>2711</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	2711	2711	N-LINKED (GLCNAC. . .)
FT <th>CARBOHYD</th> <td>2721</td> <td>2721</td> <td>N-LINKED (GLCNAC. . .)</td>	CARBOHYD	2721	2721	N-LINKED (GLCNAC. . .)
FT <th>VARSPLIC</th> <td>349</td> <td>2431</td> <td>MISSING (IN ISOFORM V3).</td>	VARSPLIC	349	2431	MISSING (IN ISOFORM V3).
FT <th>VARSPLIC</th> <td>2697</td> <td>2738</td> <td>PSAQRKSKYKLLSSSVKDNSINTSKHHRMSRRROENR</td>	VARSPLIC	2697	2738	PSAQRKSKYKLLSSSVKDNSINTSKHHRMSRRROENR
FT <th>CONFLICT</th> <td>2535</td> <td>2539</td> <td>R -&gt; RKMSPKKQPCPNKY (IN ISOFORM VINT).</td>	CONFLICT	2535	2539	R -> RKMSPKKQPCPNKY (IN ISOFORM VINT).
FT <th>SEQUENCE</th> <td>2738</td> <td>AA: 300004</td> <td>AERC -&gt; NSAR (IN REF. 4).</td>	SEQUENCE	2738	AA: 300004	AERC -> NSAR (IN REF. 4).
FT <th>SEQUENCE</th> <td>2738</td> <td>AA: 300004</td> <td>MM: 12CA626D58BDC6A CRC64;</td>	SEQUENCE	2738	AA: 300004	MM: 12CA626D58BDC6A CRC64;
Query Match				
Best Local Similarity 19.2%; Score 130; DB 1; Length 2738;				
Matches 177; Conservative 116; Mismatches 354; Indels 272; Gaps 41;				
QY	32	QNNPANIINHPAMDFAINQAKAGNPVLLTPBQIQARLNAGI-NAKPOSQALDVNFD	90	
Db	750	EDGEEDCVANADVTTTPSVQYITGKPHVTVPKNPEAAEARGLYESVAPSO-----NFS	804	
QY	91	DOSPISRIQEGOSPRLGIDMSVIEETPLSLBELFAQESTENGINPNYIPEYOGEOGNSE	150	
Db	805	NISATDT--HOFIPAETVELS---TMOFTKSKBATELLEITWKPERY-PETPEHFSGE	857	
QY	151	VVPEPTPEKPRGLIKRLYLARLFNDGCVNKKVRLAKAFYQOSOGSETSAISSHOKTEPRA	210	
Db	858	PDVFPTLPS-----HDG-KTTKMEFETTESNPNT-----NPEHOKP--	895	
QY	211	NIKALEDITQESAMDLSGIPRLQRTALVAARAVG---YDIDLSIRNSIGEVNIIH	267	
Db	896	-IPLPEBEFSEGAID-----QASQOQIFISRAPEVALGKGTDOSPTISTISINSGSVNH	949	
QY	268	DLGERVYIDIRAVNVRKSGADDKAFYTV-----ADEVPLLIQDVFNHGK	311	
Db	950	ALEED-----PIALGTISQTDSEMSSTVESVMEVTPSQTFEFGSSSAPRIEES-----GE	999	
QY	312	YETKNLLENASAHGVEFDGKWLDRSDVILLPDTAVSL-----YDTGQYQFDEVV--	365	
Db	1000	VEETNKIENFTVYDLPORE-----PYDTLLPDMSSIMITDHHIITPRATYARLDQSLPS	1053	
QY	366	-----FTTIDPKTNQLT-----TDPDK-----LPKRELLQELLVNNMG	399	
Db	1054	TDARPTQFGIQTITSEWVSSPFGKRTEDDKEDKRDYNAATITGEVQVATERSDYLITSEL	1113	
QY	400	EAYVQLQAVRAL-----SNDLIATRYFNWVTEIVPEREBOION-DQVSFEQSSSR	449	
Db	1114	ESSNVAASSPLDTWEGFVPEPTSTVSEK-EMANTTPVFETSDVANLETFOSPEHSSSS-	1170	
QY	450	TEPNOVE-----SLEP	462	
Db	1171	-QRPVQBELTTLGSKPRLIMDGLSGDASTDMEFTTASFTLDESDTGKKEKLPPTLSR	1229	

```

QY 463 VIET-----VELTOSILMDISPIEF-----SASNLQDKLNLVAKA--RHLYDMPDRV 510
Db 1230 SVETSSSSPTIGLAPSTVLDIEIVEMNOTSKTLLISELGPSTSOAEVRDLY----- 1282
QY 511 LAINHDDGVNRSLIGRISDAVSAVARAILPDE-----SENEVIDLPERTALANRKT PAD 564
Db 1283 -----PGIGEPSSGSSSE-YPTVSTTKMEELVGMGSENERV-----KD 1321
QY 565 VYQSKKVPILYVAVASDKPRDGOIGLWGSDDTGLVTKFEHNLINRQIOAGAEILRLSED 624
Db 1322 TQTLSSIP-----PTSDNINPVPDSKGFSTVASTAFPMWEEFMTSAEG--SGEE--LSSV 1373
QY 625 KKGVKYKATKPL-----SHPLNDQLATLGYQOEVEFGHSHNGRDLSTRLEHEISRS 676
Db 1374 RRSVSL--VPLPLGVDLLPTTESPFYQDEFEBAAVTBAKQKQALPLAVSGNTVDLTENRD 1431
QY 677 IIONGWMNRTYSLRYRLDKLKTQAPPEY-----QDLPVDFVNGKPSQOAL 722
Db 1432 IEVNS-----TMS-----VDLPQTEPAKLMSKPEVNPKEQIGSETVTOQKAGQKSFESL 1483
QY 723 LAGVANHKTVAD-----NLVPMRGYRQRTSLEVGSSGLVSDANMAIARAG 768
Db 1484 HSLAEQTLTLOSILTEEVOTSYSLMTTKYTNFNEVEEGTSI--AHMSTPGCP 1540
QY 769 ISGVYSGDNAYGSNRAHOMT 789
Db 1541 IKGLSEYPTHPEATGKSYSPS 1561

RESULT 28
SLAP_BACCI STANDARD: PRT: 1616 AA.
AC P35824:
DT 01-JUN-1994 (Rel. 29. Created)
DT 01-JUN-1994 (Rel. 29. Last sequence update)
DT 01-FEB-1996 (Rel. 33. Last annotation update)
DE 5-layer related protein precursor.
GN BUTB.
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1397;
RA 11)
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B3312;
RX MEDLINE=94374689; PubMed=7522196;
RA Aubert-Pivert E., Davies J.;
RT "Biosynthesis of butirosin in Bacillus circulans NRRL B3312:
RT Identification by sequence analysis and insertional mutagenesis of
RT the butb gene involved in antibiotic production.";
CC 1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. MAY PLAY A ROLE
CC IN THE EXPORT OF BUTIROSID FROM THE ORGANISM.
CC 1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L20421; AAA62588.1;
CC InterPro: IPR003443; Big_2.
CC InterPro: IPR001119; SLH.
CC Pfam: PF00395; SLH; 1.
CC Pfam: PF02368; Big_2; 1.
CC Signal: Cell wall; S-layer.
CC SIGNAL 1 30
CC CHAIN 31 1616 POTENTIAL.
CC FT S-LAYER RELATED PROTEIN.
CC SEQUENCE 1616 AA; 172874 MW; 1F03FF7A4F113AA7 CRC64;

```

```

Query Match 2.7%; Score 129.5; DB 1; Length 1616;
Best Local Similarity 17.5%; Pred. No. 27;
Matches 203; Conservative 169; Mismatches 407; Indels 379; Gaps 51;

QY 34 NPANINHPADHTALINQAK-----AGNPVLLPPEQIOARLNA----- 72
Db 189 NPDPGLMLVTASCTTYTSPDFYSGOVVRKSGQGISLTYNTEALAKADGAFQSAVRLT 248
QY 73 AGLN-----KPSQALDVNFDOSPISRIQSGSPPLGLMSVIEETPLS-LE 121
Db 249 AGTNALSVKLLKGRREIVSTVTYNDAQPADLLEVAAPADITISGPAHALGYVD 308
QY 122 ELFAQSTENGINPNQIYPIEYQGEQPNSEY-----VPEPLEPEKPKILKILYAR 171
Db 309 QDLAGIDDTVALFTNDWGPQITVPQFNVAQVDAGSKVTKVNPSPIDGKTPAWTGPTDLE 368
QY 172 LFNQGVNKKVPRLKAQFYQSSQSGSETSAIGSSHOK--TEPYANIKAALE-----DITQESA 224
Db 369 IPSGGYVLAQ-----DTSYAGKNIKKYLATYFKVGDALIKRKNQFAVVPVKDL 416
QY 225 MDINGSIPR--LRQALVAARAVGYDIDLSTIRNSIGEVYIHDLGEPVYIDYRAVEY 282
Db 417 MGTGCPILARTYTLNDYAM-----YTETKPST-----ELSGTITMDDPSKI--ALTIV 460
QY 283 RGE-----GADDKAFT--TYADEVPLLIGQVEFHGKYTEKKNLEMSAHEGYEDGR---- 332
Db 461 NGTLPFPFGDGKFKTSTYTLAEGINYLIDLVYTKGKEDQSKDLV--YSRGFSSTGKAKYL 518
QY 333 WLDRSVDVILPONTADVSLIYDTGTQYREDEVEFT-----IDPKTNQTLTDP-----DK 382
Db 519 WVDQANARKFQTDGVANANPLRTAKENGVTSVEDKGVGEVYSKSTLTGAPYSAIK 578
QY 383 LPYKR-----ELLEQL-----LTVNGEATVNLQAVRALSNLLA-----TRYEN 421
Db 579 APEKASNPDLDLQEFIRYSRELGDIVHSFNIFEGSIAENFALLDSHLMBERVYN 638
QY 422 MVTEIYPEREIOINDOVSFEQSSSTRTPEAOVDSTLEPEYELTGLIMDSPIE 481
Db 639 AADNGCIKRLRESAKQANAVAFNPSPNDEVRDFOL--KTIEVLQANDV--DGVLLDARD 685
QY 482 FSA-----SNLIQDKL-NLVAKARHLYDMPDRVLAINHDDGVNR----- 521
Db 696 NESADFSDLTKAFESFLGARKOLNMPD--VFYAGNVAKDGPLLRDMWFFRSKTI 752
QY 522 -----SLIGRISDAVSAV-----ARAILPDES- 543
Db 753 KSFTSEVRQLTDRVKAKEGKKIEVSAYVGSWFESYLVNGVHMGSTERRDERLRMKDSV 812
QY 544 -----ENEVIDLPERTA--LANRKT PADYQSKKVPILYVAVASDKPRDGOI 587
Db 813 YTPGYEESGVNKLDDIMGAQVOTTAPELHYITLONITNGSVPLAAGA----- 863
QY 588 GLWGSDDTGLVTKFEHNLINRQIOAGAEILRLSEDKKGVKILYATKPSHPL----- 640
Db 864 -----LTNVOEPALQHDVFOAGL-----ANTGILMFDASQVWMPVAGALRN 906
QY 641 -----NQQLRATLGYQOEVEFGHSHNGRDLSTRLEHEI 673
Db 907 LYYVRDYQIGISLPDPSDFLBSGIYNTNLIENNIVLTDTFSTYSGNSRFV----- 959
QY 674 SRSIIIONGM--NRTYSLRYRLDKLKTQAPPEY-----VD----- 711
Db 960 -EAYVDSSGKVTISVPKQTQAMTNMCK-----PDEINSVIIPKGFVYVTLDSAGITRKQ 1013
QY 712 -----FVNGKPSQOALLAGVAVH--KTYADNLVPMRGYRQRTSLEVGSSGLV----- 758
Db 1014 LVANAYETGDSVRAAALSGFLAYEGLRTSADSVT--FRG-----KVDVLPKQASVTYNG 1066
QY 759 -----DANMAIARAGISG-----YVSRGDA-----YG 781
Db 1067 QEALREDGTFOADVIRPGANPVYIIVRGDFKINEKTVTTIIGDEAAVAKALKLDRGYS 1126

```



QY 782 SNRAHQTGCIQAGYINSNDNFHVPYRLRFPAGDQSTIRG-ANDSLSPISDKGYLTGQ 840  
DB 1127 MNKESJRLAVTAET-----SSSSTDPVTCQAAYASLDPAVVSVDAT-GR 1169  
QY 841 VLAV----GTAENYVE-FMKDLRLAV-FGDGNAYDKGFTMDTKIGAG-VGVNNAEPVQ 893  
DB 1170 ITALRESSGVQATYEHMTATARVSVSGTGGSDTGGSGTGGSGAGGGTAPSGP 1229  
QY 894 VRVDVATGVEEGNPRL 911  
DB 1230 ERTSVETKDSGRNLT 1247

RESULT 29  
FRPA.NEIMC  
ID FRPA.NEIMC STANDARD: PRT; 1115 AA.  
AC P55126;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Iron-regulated protein frpa.  
GN FRPA.  
OS Neisseria meningitidis (serogroup C).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxId=135720;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-FAM20 / Serogroup C;  
RX MEDLINE-93139051; PubMed-8423153;  
RA Thompson S.A., Wang L.L., West A., Sparling P.F.;  
RT "Neisseria meningitidis produces iron-regulated proteins related to  
RL the RTX family of exoproteins.";  
J. Bacteriol. 175:811-818(1993).  
CC -!- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL  
CC DISEASE.  
CC -!- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED.  
CC -!- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING  
CC CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC  
CC ACTIVITY.  
CC -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL: L06302; AAA25454.1;  
DR InterPro: IPR001343; Hemlysn\_Ca\_bind.  
DR Pfam: PF00353; hemolysincbind; 7.  
DR PRINTS: PR00313; CABDNCRPT.  
DR PROSITE: PS00330; HEMOLYSIN\_CALCIUM; 5.  
KW Toxin; Calcium; Outer membrane; Repeat.  
FT DOMAIN 765 988 13 X REPEATS, GLY-RICH.  
FT REPEAT 774 779 2.  
FT REPEAT 783 788 3.  
FT REPEAT 902 907 4.  
FT REPEAT 911 916 5.  
FT REPEAT 920 925 6.  
FT REPEAT 929 934 7.  
FT REPEAT 938 943 8.  
FT REPEAT 947 952 9.  
FT REPEAT 956 961 10.  
FT REPEAT 965 970 11.  
FT REPEAT 974 979 12.  
FT REPEAT 983 988 13.  
SQ SEQUENCE 1115 AA; 122177 MW; 21841065BCCACEFA CRC64;

Query Match 2.7%; Score 129; DB 1; Length 1115;  
Best Local Similarity 19.9%; Pred. No. 17;  
Matches 205; Conservative 102; Mismatches 335; Indels 388; Gaps 45;

QY 56 NPVLLTPEQIQARLMAAGLNAKPOSQALDVYVNDQSPISRGQSPPLDMSVIEET 115  
DB 155 NSRMKLSSTVQLQHEHNLVIGFKIK-----DYN-----ERLGE-----SIMIDDF 194  
QY 116 TPLSLLELFAQESTEMGINPDYIPEYQGEQPNRNVVPPPLPEPKP 163  
DB 195 TPKSIANFEA-----DPTYSNVLEEVSRFTYSLVPPDANPWPKGEDYIGRGISEWG 246  
QY 164 -LIRLYARLEFNDGVNKKVPRKAEYQ-----SSQGETSAIGSSHQKEPYANIKAALED 218  
DB 247 ELLEKWKQDF-----LPYLEKEMDQPKFREDMLPEPPEARBMKIDPKRSKHYHYD 300  
QY 219 ITQESADNLGSIPLRKLQALVAA--AVGYIDIDLSIIRNSICEVD---VIHIDGEP 272  
DB 301 ---PLALDLDDG-----GIETVAAKGFGALFDHRNOGIRFATGVASADDLTLVDLNGN 352  
QY 273 VYIDYRAVEVREGADDKAFTTVADEVPLLIGVFHHGKYETKKNLLENASAHEGYFDGR 332  
DB 353 GLID-----NGAE-----LEGD-----NTK--LADGSPAKHGVALA 382  
QY 333 WDRSVDVILPDNTADVSL-----IYDTGTQYRFDEVEVFTID----- 370  
DB 383 ELDSNGDNIT--NAADAAPQLRWQDNLNDGISOANELRTLEELGQSIDLAKDVNKN 440  
QY 371 -----PKTNQLTTD-----PKLIPVRELLEQLLVNMGDAY 402  
DB 441 LGNGNTLAQGSYTKTGTAKMGDLLAADNLHSRFDKVELTAEDQAAKANLAGIRLR 500  
QY 403 NLAQVRLSNDLILATRFVFNVNTEIVEPEPEQIONDVSFEQSSSTEPRAVDSESTLEP 462  
DB 501 DLREAAALSGDLA-----NML-----KASAAETKRAQL--ALLDN 534  
QY 463 VIETVELTDGILNDISPIEESA-----SNLIODKLNLVAA--KARHLYDM 505  
DB 535 LIHKMAETDSNMKKSPKMRSTDMTOTANEGIALTPSQVQALKNALVSLDKRAAIDA 594  
QY 506 PDDRVLAINHDDGVNNSILGRISDAVASAVARAILPDESENEVIDLPRTALANKRPADY 565  
DB 595 ARDRIAVLDVATGQDSSSTLYMS-----BEDALNIVKVINDT 631  
QY 566 YQSKKVPY---VFVASDKRRDQIGLGMSDTGT---RLVTKFEH-----NLINR 610  
DB 632 YDLAKNITQNLLEQYRLQPLYNQISFKMENDEFTDFSGLYQAFNHVKEPNQKAFVDL 691  
QY 611 DGYQAGAEUR-----LSEDK----- 625  
DB 692 AEMLAAGELRSWYEGRRIMADYVEEAKKAGKFEQYKVLQGFVALLAKTSGTQADDILQ 751  
QY 626 -----KGKLYATRP---LSHPLNDQLRATLGYQOEYFGHS----- 658  
DB 752 NVGFGHNKNVSLYGNNDGNDPLTIGGAGNDYLEGSSGSDTYFGKGFQDPTVYNNDYATGKR 811  
QY 659 -----TNGF--DLSTLTLE--HEISR-----SIIONGWNRTYSLARYDLKLT 698  
DB 812 DIIRFTDGTITAMLFTTRGHNHLIRAKDDSGQVTVQSYFQNDG-----SGAARIDE--- 863  
QY 699 QAPRETWQDLPYDFVNGKRSQSEALLAGVAHVKTVAQNLNPNMNGYRORYSLEVG--SGI 756  
DB 864 -----IHDNGKLVDA-----TVKELVQOSTDGSRLYAYOGSGTLNGG 903  
QY 757 VSDANMALIARAGISGYVSEGDNAVGSNRAHOMTGGIOAGYIMSDNFHVPYRLRFPAGD 816  
DB 904 LGD-----DYLIGADGNDLLNGAGDNGNSIYSGN-----GND 934  
QY 817 QSIRGAHDSLSPISSDKGLTGGVYLAAGTAEYNTPEMKDLRLAVFGDIGNAY-DKGFVN 875  
DB 935 TLDDGEGNDALTYGNGDNLNGE-----GNDHINDEGND--TLIGAGNDYLEGSSGS 987  
QY 876 DTKI-GAGVG 884  
DB 988 DTYVEGEGCG 997



```
Db 762 SNKPLEDLEPEKOVEAKSLSEKTSVATVA-----GGAVVGATKTHSATKASTA 812
Qy 631 YATKPLSHPLNDOLRALIYGQOEYFGHSTN-----GFDSTRT-----LEHE 672
Db 813 ASKSKSTETL--VMKKTASTSTSYGANKSAAPRPARLGIKSTATITSTTSLTGN 870
Qy 673 ISSSIITONGGWNRTYSLRYLDLKTQAPPEWODLPVDFVNGKPSOEALLAGVAHRTV 732
Db 871 PRKSLSSMWG--STVKPPTLSTGRPATAPVSKVTLGAKITNKPTASG-----TA 919
Qy 733 ADL-----VAPMAGYRQRYSLFVSSGLVSDANMAIRAGISGYRSGDNAYGS 782
Db 920 SDNVYTRTTLPLVSTNARPRATSGTGSVAS--STARRPVT-----NAKGS 962

RESULT 31
DPOL_BACSU          STANDARD:          PRT:          880 AA.
AC 034996;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA polymerase I (EC 2.7.7.7) (POL I).
GN POLA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=1423;
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Betero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Broillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Ertlan K.D., Erlington J., Fabret C., Ferrari E., Foubert D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,
RA Gilm S.T., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepti G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue Y.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Portollik S., Prescott A.M.,
RA Preecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche S., Schreiber S., Scoffone F.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Sadale Y.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serrot P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosto V., Uchiyama S., Vandenbol M., Vanlier F., Vassarotti A.,
RA Viati A., Wambuit R., Wedler E., Wedler H., Welternegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -I- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY (BY
```

```
CC SIMILARITY).
CC -I- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}(N).
CC -I- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS.
CC -I- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-A FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF008220; AAC00350.1; -.
CC EMBL: Z99118; CAB14869.1; -.
CC HSSP: P52026; 1XWL.
CC Subtilist; BG12656; POLA.
CC Interpro: IPR002562; 3_5_exonuclease.
CC Interpro: IPR002421; 5_3_exonuclease.
CC Interpro: IPR001098; DNA_pol.
CC Interpro: IPR002298; DNA_pol.
CC Interpro: IPR000513; EXO_N_I.
CC Interpro: IPR003583; HHH_1.
CC Interpro: IPR003584; HHH_2.
CC Pfam: PF00476; DNA_pol_A; 1.
CC Pfam: PF01367; 5_3_exonuclease; 1.
CC Pfam: PF02739; 5_3_exonuc_N; 1.
CC PRINTS: PR00868; DNAPOLI.
CC SMART: SM00474; 3SEXOC; 1.
CC SMART: SM00475; 53EXOC; 1.
CC SMART: SM00278; Hhh1; 1.
CC SMART: SM00279; Hhh2; 1.
CC SMART: SM00482; POLAC; 1.
CC TIGRfams: TIGR00593; polA; 1.
CC PROSITE: PS00447; DNA_POLYMERASE_A; 1.
CC Transferrase; DNA-directed DNA polymerase; DNA replication; DNA repair;
CC Hydroxylase; Exonuclease; DNA-binding; Complete proteome.
CC SEQUENCE 880 AA; 99091 MW; AFB434ARDF26FA48 CRC64;
SQ

Query Match 2.7%; Score 128.5; DB 1; Length 880;
Best Local Similarity 17.3%; Pred. No. 12;
Matches 174; Conservative 170; Mismatches 355; Indels 307; Gaps 46;

Qy 8 ANRSEFMPVAL-----AAVLPMTSQLAQNNPANTINHPAHD-----ALNQA 52
Db 16 AYRAFEPFLPLISNDKGYHTNAVGFAMITLKMLEDKPTHTLAVFDGKTTFRGTRKEY 75
Qy 53 KAG---NPVVLTPPEQ---IQARLNAAGLNAKPQSQALDVVNFDDQSPISRIQSPPLG 106
Db 76 KCGNQKTPPEL---SEQMPTFIRELDAY-----QISRYELEQYFADDIIGTLAKSAEKDG 127
Qy 107 LDMSVI---EETPPLSLIEELPAOESTEMGINPND-YIPEYQGEQPNSEVVVPTLPEKRP 162
Db 128 FEYKVFESGDDLIQLATDKT-TVAITRKGITDVEFYFPEHKEKYG-----LTPEQI 178
Qy 163 GLIKRLYARLNFQNVNVPPLAKKFYSSQSGERSAIGSSHQKREPRANIKAALEDITOE 222
Db 179 IDMK-----GLMGSSQNDIPVPG-----VGEKTAI-----KLKQPDSEVKLESIDEV 223
Qy 223 SAMDLNGSIRLROTALVARAVGYDIDISITRNSIGEVDVITHDGEPEYIDYRAVEV 282
Db 224 SGKKLKAKKLEEFQDALMSK-----ELATIMTD--APIEVSQSLLEY 263
Qy 283 RGEQAD-----DKAFTVADEVPLLIGVFNHGGYETKKNLIENASABHGFDFGRWLD 335
Db 264 QGFNRQEVIAIFDGLGHNTLLER---LGE-----DSAEAQ-----D 297
Qy 336 RSYDVI---LPDNTADV-----SLIYDGTQYRFDEYVFETIDPTNQLTDPDLKPV 385
Db 298 QSLIEDIVKTVTDVTSILVSPSAFVEQIGDNTHEPILGFSI----- 341
Qy 386 KRELLEOLLTVNNGEAVNIQAVRALSNDLIATRYFNNVNTIEIVPEREQIQNDQVSEQGS 445
```

```
Db 342 -----VNETGAFYFKDAVESEVF-----:|:|:|:
Oy 446 SSSRTTBA-----QVDESTLEPIETVELTDGLMDISPIEFSASNLIOOKLMLVAKA 499
Db 376 DSKRAVVALRMQGIETLGAEPDTLLAIIINPG-----NSYDDVASVADYG 422
Oy 500 RHLVMDPDRVLAINHDDGVNRSI-----LGRISDAVSAVARAILPDESENEVDL 550
Db 423 LHITS-SDESV-----YGGAKRAVPSEVDLSEHIGRAKALAIQSIREKLVQLENNDOJEL 477
Oy 551 PERTALANRKTADVYOSKVPYLVFAVADSKPRDGOIGLGSGSTGTRLVTKFEHNLINR 610
Db 478 -----FELEMPPLALI-----LGEMESTG-----KVVVDRIKR 506
Oy 611 DGYOAGAEIRLSEDKKGVATKPLSHPLNDOLRATL-----GYOQVEFGSTNGFD 663
Db 507 MGEELGAKKKEERK-----IHEIAGEPFINSPKOLGVLIFERIGLPVYKKTGTSTADV 564
Oy 664 LSTRLEHEISRTIION--GGMNRTY-----SLRYRLDKLKTQ-----APP 702
Db 565 LEKLADKHDIYDIYLOQRQIKLQSTYIEGLKVTYRPSDKHVHTRFNOALTQGTGLSSTD 624
Oy 703 ETWQDLPVDVFNKGRSQEALLAG-----VAHKTYADNLV-----NPMR 741
Db 625 PNLCNITPIREEGKRIKQAFVPSKDWLIFADYSQIELRVLAHISKDENLIEAFTNDM- 683
Oy 742 GYRORYSLEV--GSSGLVSDANMAIARAGISGVYSPGDNAGSNRAHOMTGGIOAGYIWS 799
Db 684 DIHTKTAMDVYHAKDEYTSAMRQAKAVNFGIV-YGISDYGLSQNGITKEKAGAFI-- 740
Oy 800 DNEHNVYRLRFFPAGGDSITGYAHDLSPLSDGYLGTGGVAVGTALEYNYEFPKDLRL 859
Db 741 DRY-----LESF-----QGVAYMEDSVQAKQGYVY-----TLMHRRYIPELTS 782
Oy 860 AVFEDIGNAYDKGFTNDPTKIGAGVGRWASPVQGVNVDATGVKEE 905
Db 783 RNFN-----IRSFARERAMNPTIOGSAADILIKAMIDMAKLEK 822

RESULT 32
DSG3 HUMAN
ID DSG3 HUMAN STANDARD: PRT: 999 AA.
AC P32926:
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Desmoglein 3 precursor (130 kDa pemphigus vulgaris antigen) (PVA).
GN DSG3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=92069753; PubMed=1720352;
RA Amagai M., Klaus-Kovtun V., Stanley J.R.;
RT "Autoantibodies against a novel epithelial cadherin in pemphigus
RT vulgaris, a disease of cell adhesion."
RL Cell 67:869-877(1991).
CC -|- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
CC -|- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND
CC CARCINOMAS.
CC -|- DOMAIN: CALCITONIN MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -|- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN
CC DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE
CC LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES
CC AGAINST DSG3.
CC -|- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M76482; AAA60230.1; -.
DR PIR: A41088; ITHUG3.
DR HSRP: P15116; INCU.
DR Genew: HGNC:3050; DSG3.
DR MIM: 169615; -.
DR InterPro: IPR002126; Cadherin.
DR Pfam: PF00028; cadherin; 4.
DR PRINTS: PR00205; CADHERIN.
DR SMART: SM00112; CA: 4.
DR PROSITE: PS00232; CADHERIN_1; 3.
DR PROSITE: PS0268; CADHERIN_2; 4.
DR Cell adhesion: Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.
FT SIGNAL 1 23
FT PROPEP 24 49
FT CHAIN 50 999
FT DOMAIN 50 615
FT DOMAIN 616 640
FT TRANSMEM 641 999
FT DOMAIN 159 268
FT DOMAIN 269 383
FT DOMAIN 386 499
FT REPEAT 910 935
FT REPEAT 936 966
FT CARBOHYD 110 110
FT CARBOHYD 180 180
FT CARBOHYD 459 459
FT CARBOHYD 545 545
FT SEQUENCE 999 AA; 107503 MW; 60479DD6AC219A1 CRC64;

Query Match 2.7%; Score 128.5; DB 1; Length 999;
Best Local Similarity 19.9%; Pred. No. 15;
Matches 185; Conservative 129; Mismatches 377; Indels 239; Gaps 46;

Oy 95 ISRGESPPGLDMSVLEETPLSLLEFLAQESTEMKINPDYIP-----EYGGQSPRSE 150
Db 87 ISGVGIDDPPEGI-FVVDKNTGDNITAIVDKEET-----PSFLITCRALNAQG----- 134
Oy 151 VVVPPTLEPERKGLIKRLYARLFNDG--VNKVPRLAKAFYOSOSGETSATGSSHQKTEP 208
Db 135 -----LDVEKP-LITVTKIINDNPNPVFSQIIFMGEIEENSASNSLWMLNATDADPE 187
Oy 209 -YANIKAALEDITQESAMDNLGSIPLRLQOTALVARAVGYDIDISTIRNSIGEVDYIHH 267
Db 188 NHLNSKIAFKITVSOEPA-----GTPWFLLSRNTGVEVRLTN 223
Oy 268 DLGEVYIYDAVEYRGSGADKAFIT-----VADEVPLLIGVFNHGKYEK-- 315
Db 224 SLDRQASSTLYV-VSGADKQEGISTQCECNIAKKVDYNDNPFM-----FRDSQISATIE 277
Oy 316 KNLENSAEGYEDGRWLDNRSDVDILPDNTADVSLIYDTGTQYRFEDEVVEFTIDPKTNQ 375
Db 278 ENILSSELL-----RQVYTDLDEYTDNMLAV-YFTSGNKGWMEF-----QTPDPRTNE 326
Oy 376 LTTDPDKLPVKREL-LBDLLTYNMGEAATNLQAVRLASDLATRFNNVNTIYVPEREQ 434
Db 327 -----GILKVVAKALDYEDQLQSVKLSIAVK---NKAEPHQSVISRY-RVQSPVVTI-----Q 373
Oy 435 ION--DOVSFQSSSSRREPQAVDESTLEPIYEVVELTDGLMDISPIEFSASNLIOOKL 492
Db 374 VINRREGIAFRPAKST-----FTYQKGISSKLVDIYL-----GYQAIDDDT 416
Oy 493 NLVAKARHL-----YDMPDRVLAINHDDGVNRSILGRISDAVSAVARAI----- 538
```

```

Db 417 NKAASNVKVRGNDGGTGLMTDSTAEIKFKNNMRDSTFLVNTKTITAEVALIDEYTKT 476
QY 539 -----LPDESENEVIDLPETALANRKTADYQOSKV-----PLYEVASDKPRD 584
Db 477 STGVYVVRVPFNNCPFAVEKDAVCS-SSPSVVSARTLNRTGP-YTFALDDQ-- 532
QY 585 GQIDLGWSDT--GTRLVTK-----FEHNLNRDGYOAGAEIRLSE-----DKK 626
Db 533 VKLPVAVSITLNTATSAALRAEOEIPGVYHISLVLTDSQNNRCMPRSLFLEVQCQNR 592
QY 627 GVKLYATKPLSHP-----LNDOLKATLGYQDEVFHGS 658
Db 593 GI-CGTSTPTTSPGTRGRPHSGRLGPAIGLLLLGILLLLAPLLLTTCOGASTGV 651
QY 659 TNGF-----DLSTRLEHISRSIIIONGMNRTYSLRYRLDKTKTO--APPETWDLPDF 712
Db 652 TCGTIPVPDSEGTI-HQ-----WG--IEGAHPDEKEITNICVPPVT----- 690
QY 713 VNGKPSOEA-----LLAGVAHVHTYADNLVNPGRYRORYSLEVGSSGLVSDANMAIA 765
Db 691 ANGADFESESEVCTNTYARGTAVEGTSGEMETTKLGATESGAAGFATGVGSAASGFG 750
QY 766 RAGISGVYSPEDNAVYGSRAHQMTGGIOAGY---IWSDNFNHVPRRLFFRAGDOSINGY 822
Db 751 AATGVGICSSGQS--GTMRTRHSTGTNKKDYADGAISMNFDLSYFSOKAFACAEEDDQOE 808
QY 823 AHDLSLPSDKGY-LTGGQVLAAGTAEYNEFMKDLRLAVFGDIGNAVDKFTNTKIGA 881
Db 809 ANCDLLIYDNGADATSPGVSVCSCSTIADDDSLDLSIG-----PKFKLAETSL 861
QY 882 GVGVRMASPVGVQVRVAVATGVKEGKNPKL 911
Db 862 GVDGE-GKEVOPSPKDSGYGIESCGHPLEV 890

RESULT 33
RBP1_PLAVB STANDARD; PRT; 2869 AA.
AC 000798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN RBP1.
OS Plasmodium vivax (strain Belém).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites.";
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
CC EMBL: M8097; AAA29743.1;
KW Malaria; Receptor; Signal; Transmembrane.
FT CHAIN 1 2869
FT SIGNAL 17
FT DOMAIN 18 2807 EXTRACELLULAR.

```

```

FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EB0FF CRC64;

Query Match 2.7%; Score 128.5; DB 1; Length 2869;
Best Local Similarity 18.9%; Pred. No. 75;
Matches 149; Conservative 127; Mismatches 302; Indels 209; Gaps 36;

QY 15 VALAIVLPLM-TSQL-----AQQNNPANIINHPHDAIN-----QAKAGNPVLLTP 63
Db 2180 ISLISYTANMKTSKRLIMINKENKTEKVCVDYIKNSSSTDGVETLKGFGYSKLTFFSSA 2239
QY 64 EOIQARLNAAGIN-AKPOSAIDVNPDDQSPISNIGESPPGLGDMVIE--ETTPSL 120
Db 2240 SETVONADPTYSVNFKKHEKESLNAIR-DIKKELYLFHQSS-----DISTVEGVONMLNL 2293
QY 121 BELFAQESTEMGINPNDIYIPEYQGEQPNSEVVPPLEPEKPKGLIKRLYARLFNDGVNKV 180
Db 2294 YDKLINEERKREMDLRYNISETRKQMEHSTDFKRMIELHK-----MNETNKS 2343
QY 181 PLKAKFTYQSSQSGETSAIGSHQKTEPYANIKAALEDITQESAMDNLNGSIPRLROTALV 240
Db 2344 LLEKEKKLKS-----NDHMSMEAEWIKNGLK-YTPESYONINNI----- 2383
QY 241 AARAVGYDIDLSIRNSIGEVDYIHDGEPVYIDYRAVEVRGSGADKATTVTADEV 300
Db 2384 -----YSVEAEVKLTLEIDROYG-----NYQIVE-----EHKKQSTILIDRTN 2423
QY 301 LLIGDVFHNGRYETKKNLE-NASAEHGYFDGRWLDLRSDVLLPNTADVSL-----YD 354
Db 2424 ALMDOI-BIFKKNYNNLMEVNTETIHRVND--YIEKTKNLKYQAKTEYEOLENIKQND 2480
QY 355 TGTQTRFDEVV-----FTIDPKTNQLTTD--PDKL-----PVKRELLEQLLTVM 398
Db 2481 DMLQNIFFLKVSIIEFENVKKKESIINDLYEQERLLKIGEHLEIKRNVETL----- 2535
QY 399 GEAVNL-QAVVALSNDLATRYFNVNVEIYFPERE--QIQNDQVSFQOSSSKRPEAOV 455
Db 2536 -SSTEIDQKMEMSKNLEKSKMNNTYSIYLEEAEININDAKQIKDDT-----I 2587
QY 456 DESTLEPIETVELTDGILMDIS-----PIEFSASNLIDOKLNLV-----AKARHLUDMPD 507
Db 2588 LNSVLEALQKRGMDALFQSMADRNPNKYSAKYNNANEIIRQLEVKLREIGOLVQ 2647
QY 508 DRVLAINHDDVNSILGRISDAVASV-----ARAILPDESENEVIDLPETALANRKT 562
Db 2648 D-----SESIISEMSSKKSATIEKERTARL--RTSENNRREERERARQEMSMN 2694
QY 563 ADVYQSKVPLVYVVASDKPRDQIGLGWSDTGTRLVTKFEHNLNRDGYOAGAEIRLS 622
Db 2695 NDPQSEPT-----HSEGSIGEGKESDSDETGLT-----HDAGADEDST 2733
QY 623 EDKGVKLYATKPLSHPLNDQLRATLGYQGVFGSTNGFPLSTRLEHISRSIIIONG 682
Db 2734 SSAG-----AHELEEETTAPEMETBMNDTLGLTY-TTRSDPDMHTENTQCG- 2782
QY 683 WNRITYSLRYRLDKLKTQAPPETWDL-----PVDFVNGKPS-----QEALL 723
Db 2783 -----TYQDTSNSDEADILNKKFNNVVKYAGAFVLLCTSAVI 2821
QY 724 AGVAVHK 730
Db 2822 GATIAHK 2828

RESULT 34
FRPA_NEIMB
ID FRPA_NEIMB STANDARD; PRT; 1302 AA.
AC G9K0K9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Iron-regulated protein frpA.
GN FRPA OR NMB0585.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.R.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Uitterlbeck T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scalapato V., Maignant V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RL MC58";
RU Science 287:1809-1815(2000).
CC -1- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL
CC DISEASE.
CC -1- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED (BY
CC SIMILARITY).
CC -1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CC CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
CC ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE002414; AF41013.1; -
CC TIGR: NMB0585; -
CC InterPro: IPR001343; HemIysn_Ca_bind.
CC Pfam: PF00353; hemolysinCbind; 12.
CC DR PRINTS: PR00313; CABDNGRP.
CC DR PROSITE: PS00330; HEMOLYSIN_CALCIIUM; 9.
CC KM Toxin; Calcium: Outer membrane; Repeat; Complete proteome.
CC FT DOMAIN 752 975 13 X REPEATS, GLY-RICH.
CC REPEAT 761 766 2.
CC FT REPEAT 770 775 3.
CC FT REPEAT 789 894 4.
CC FT REPEAT 898 903 5.
CC FT REPEAT 907 912 6.
CC FT REPEAT 916 921 7.
CC FT REPEAT 925 930 8.
CC FT REPEAT 934 939 9.
CC FT REPEAT 943 948 10.
CC FT REPEAT 952 957 11.
CC FT REPEAT 961 966 12.
CC FT REPEAT 970 975 13.
CC SO SEQUENCE 1302 AA; 141397 MW; 21D058C56C98BDE8 CRC64;
Oy Query Match 2.7%; Score 126.5; DB 1; Length 1302;
Db Best Local Similarity 19.6%; Pred. NO. 29;
Matches 184; Conservative 92; Mismatches 306; Indels 355; Gaps 41;
Oy 163 GLIKRLVRLFN---DGNKVPRLAKFYOSSGSETSAIGSSHOKEPEYA-----NIK 213
Db 223 GIINDLYKSVKREWTGIFEIVNNNIKQFRDLFPNPEGWIDGHCQFAFWVKEFKRNCK 282
Oy 214 AALDELTCESAMDLDGSIPLRLQRTALVAVARAAGY---YDIDLSIIRNIGVD---YI 265
Db 283 YHYVD---PLADLDGDD-----GIEFYAAKGFSGSLFDHTNNIGIRIATGWSADGCLL 332
Oy 266 IHDLGEPVYIDYRAVEVKGADKAFETTVADEVPLLIGDVFHHGKYETKKNLLENASAE 325

```

```

Db 333 VRDLNGNGIID-----NGAE-----LFGD-----NTR--LADGSFAK 362
Oy 326 HGYPDGRWLDREVDVLTLPNTADVSL-----IYDTGTYRREDEVYFFIID-- 370
Db 363 HGYYALAEIDLSNGDNI--NAADAAFOSLRWODLNQDGIQANELRTLEELIGOSLDA 420
Oy 371 -----PKTNOLTTD-----PDKLPVKRELLEROLL 395
Db 421 YKDVKNKLNGLNGNTLAAQGSYTKTDGTAKMGDLLAADNLHSFKKVKELTAEQAANL 480
Oy 396 VMKGAEYNIQAVRAISNDLIATRYFMVNTETVFPREDOIQDQVSFEQSSSSRTPEAY 455
Db 481 AIGIRLRDIREAAALSGDLA-----NML-----KAYSAAETKEAOL 516
Oy 456 DSTLEPVEIETVELDPTGLMDISPIEFA-----SNLIQKLNLVAA---K 498
Db 517 --ALDNLHKKAEIDLSNMGKSPMLSTDWTQTANEGIALTPSQVAQLKKNALVSLSDK 574
Oy 499 ARHLYDMPDRVLAINHDDGVNRSILGRISDAVASAARAILPDESENEYIDLPERTALAN 558
Db 575 AKAAIDAAARDRIALVDAYTG-----QDSNTLYMSEEDALNI 611
Oy 559 KRTPRADVYOSKKVPLV---VPVASDKPRDGOIGLGSGDT----- 595
Db 612 VVYNTDVTYDLAKNIYQNLLEFQTRLOPYLNQISFKMENDTFLDSSGLVOAENHVAETNP 671
Oy 596 -----GTRLYVT-----KFEHN-----LINRDGY 613
Db 672 OKAFVDLAEMLAVGELRSWIEGRILMTDVEEAKKAKFEDYQKVLGOETVALLAATSGT 731
Oy 614 QAGAEIRL--SEDKKGVKIYATKP---LSHPLNDOLRATLGYQDEVFGHS----- 658
Db 732 QADDILONGVFGHNKNVSYLNGNDTLIGAGNDYLBEGSGSDTYVFGEGQDTVYNY 791
Oy 659 -----TNGF--DLSTPTE--HEISR-----SIIONGMWRTSLRY 691
Db 792 DYATGKKDILRFTDGTITADMLFTTREGNHLIKAKGSGQVYQSTFQNDG-----SGAY 846
Oy 692 RLDKLTQAPPETWODLPVDFVNGKPSQBALLAGVAHKTVAADNLVNPKRGYRQRTSLEY 751
Db 847 RIDE-----IHFDMGKVLDA-----TKRELVOQSTIDGSDRLYAYQS 883
Oy 752 GS--SGLVSDAMKATARAIGSYISFGDNAGSNGRAHQMTGGIOAQYIYSDNHNHPRYL 809
Db 884 GMTLNGGLD-----DYLXGADGDDLNGDAGNDSTVSGN----- 918
Oy 810 REFAGGDQSIIRGYAHDLSLSPDKGYLTGGQVLAAGTAENYEFMKDLRLAVFGDIGNAV 869
Db 919 -----GNDITLDGEGENDALGYNGNDALNGE-----GNDHLNGBDND---TLIGAGNDY 967
Oy 870 -DKGFTNDTKT--GAGYGVWASFPVGOVRYDVATGYKE 904
Db 968 LEGSGSGSDTYVFGKFRG-----QDAVYNYDVATGRMD 999
Oy RESULT 35
Db BCC2_ACEXY STANDARD; PRT; 1326 AA.
AC 082861;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose synthase operon protein C precursor.
GN BCSC.
OS Acetobacter xylinus.
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_TaxID=28448;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=BPR 2001;
RX MEDLINE=98296257; PubMed=9630539;

```

RA Nakai T., Moriya A., Tonouchi N., Tsuchida T., Yoshinaga F.,  
 RA Horinouchi S., Sone Y., Mori H., Sakai F., Hayashi T.,  
 RT "Control of expression by the cellulose synthase (bcsA) promoter  
 RL region from Acetobacter xylinum BPR 2001.";  
 CC Gene 213.93-100(1998).  
 CC -1- FUNCTION: Required for maximal bacterial cellulose synthesis. It  
 CC may be involved in the formation of a membrane complex for  
 CC extrusion of the cellulose product (by similarity).  
 CC -1- PATHWAY: Bacterial cellulose biosynthesis.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE ACSC/BCSC FAMILY.  
 CC -1- SIMILARITY: CONTAINS 6 TPR REPEATS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AB010645; BAA31465.1; -  
 CC InterPro: IPR003921; CellSynth\_C.  
 CC InterPro: IPR001440; TPR.  
 CC Pfam: PF00515; TPR.3.  
 DR PRINTS: PR01441; CELLSTMTASEC.  
 KW Cellulose biosynthesis; Repeat; TPR repeat; Membrane;  
 KW Outer membrane; Signal.  
 FT SIGNAL 1 30 POTENTIAL.  
 FT CHAIN 31 1326 CELLULOSE SYNTHASE OPERON PROTEIN C.  
 FT REPEAT 49 82 TPR 1.  
 FT REPEAT 291 324 TPR 2.  
 FT REPEAT 325 358 TPR 3.  
 FT REPEAT 405 438 TPR 4.  
 FT REPEAT 557 590 TPR 5.  
 FT REPEAT 701 734 TPR 6.  
 SQ SEQUENCE 1326 AA: 142103 MW: FCAB2250C4E6C4A9 CRC64;  
 Query Match 2.7%; Score 126.5; DB 1; Length 1326;  
 Best Local Similarity 19.0%; Pred. No. 30;  
 Matches 199; Conservative 134; Mismatches 344; Indels 373; Gaps 52;

QY 506 -----PDDRVLAT-----NHDDGVNRSILGRISDAVSAVARAILPDSENEVIDLPERT 554  
 DB 536 LANPVTAODRQAGILYTGSGNDAMTROLAGLSPADYSPAIRSIAEMEIKQ-DLASRL 594  
 QY 555 ALANKRTP-----ADYOSK-----KVLPLYVVASCKPRD----- 584  
 DB 595 SMVSNPVLIREALSQDPPTGARGVAVALDFRQGDVHARMAIRIASTRIDLSPPDRL 654  
 QY 585 -----GOIGLWGSDDTGR-----VTRKHNLIIND 611  
 DB 655 SYATEYMTISNPVAAARLAPLGDGTSGAGNALLPEQMOTLQOLRMGISVAQSDLLNOR 714  
 QY 612 GYQAGA-----ELRLSEDKKGVKLYAT-----KPSHPPLNDQLRATL 648  
 DB 715 GDQAGAYHLPALQADEATSPKALRLVNGHGRPKKALEIDLAVLRHNPQDDARQA 774  
 QY 649 GYQGVFEGHSTNGEDLSRTRLEHISRSIT IONGN-----NRTYS 688  
 DB 775 AVQAAY-----NSNHSNLSATRLAMDGVQESPMDARAWLMAVADQADHGQRTIEDLRAYD 831  
 QY 689 LRYRLDKL-----KTQAPPET-----WQDLFVDFVNGKPSQEA--- 721  
 DB 832 L-RLOQVEGTRAAAGPYGAHEALAPSTNPFOSRGYGHQVELGAPYTGGSYSAEASP 889  
 QY 722 -----LLAGVA--VHKTVADNL-----VNPMDRGYRORYSLEVGSSGLVSDANMAIA-----R 766  
 DB 890 DTSQMLSTINGOIR-TIRENLAPSIDGLGFRSR-SEHG-MGLUTERANIPYVRLPLQ 946  
 QY 767 AGISGVSYFGDNVAGSNRAHOMTGGIOAGYIWSDFN-----HVPYRLFRFAGGDSIRG 821  
 DB 947 AGASA-----LTFESITPTMWSGLNTGSYVDV-RYGFEMATQANQC 989  
 QY 822 YAHOSLSPISDKGYLTGS--QVLANGTAIYNEEFKKDLRLAVFGDIGNAYDKGFINDIKI 879  
 DB 990 AGHSSCGGLD--FLSANHTRIAGAGE-----AGFAPDVQF 1024  
 QY 880 GAGVGVRW-----ASPVGOVRVDVATGVK 903  
 DB 1025 GNS-----WVRADVCAASPIGFPTITNVLGVE 1050

RESULT 36  
 FRPC\_NEIMB STANDARD; PRT; 1829 AA.  
 ID FRPC\_NEIMB  
 AC 09JYVS:  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Iron-regulated protein frpc.  
 GN FRPC OR NM01415  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_Taxid=491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / serogroup B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tetteilin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eelsen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Clecio A., Parksey D.S., Blair E., Cifton H., Clark E.B.,  
 RA Cotton M.D., Uterback T.R., Knorr H., Qin H., Yamathavan J.,  
 RA Gill J., Scarlato V., Maignani V., Piza M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 RL MC58.";  
 RL Science 287:1809-1815(2000).  
 CC -1- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL  
 CC DISEASE.  
 CC -1- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED (BY  
 CC SIMILARITY).

-1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING  
CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC  
CC ACTIVITY.  
CC -1- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: AE002490; AAF41776.1; -  
DR TIGR: NM81415; -  
DR InterPro: IPR001343; Hemolysn\_Ca\_bind.  
DR Pfam: PF00353; hemolysincabind; 22.  
DR PRINTS: PR00313; CARNONGRPT.  
DR PROSITE: PS00330; HEMOLYSIN\_CALCIIUM; 17.  
KW Toxin; Calcium; Outer membrane; Repeat; Complete proteome.  
FT DOMAIN 879 1702 43 X REPEATS, GLY-RICH.  
FT REPEAT 888 884 1.  
FT REPEAT 897 902 2.  
FT REPEAT 897 902 3.  
FT REPEAT 1016 1021 4.  
FT REPEAT 1025 1030 5.  
FT REPEAT 1034 1039 6.  
FT REPEAT 1043 1048 7.  
FT REPEAT 1052 1057 8.  
FT REPEAT 1061 1066 9.  
FT REPEAT 1070 1075 10.  
FT REPEAT 1079 1084 11.  
FT REPEAT 1088 1093 12.  
FT REPEAT 1097 1102 13.  
FT REPEAT 1216 1221 14.  
FT REPEAT 1225 1230 15.  
FT REPEAT 1234 1239 16.  
FT REPEAT 1243 1248 17.  
FT REPEAT 1252 1257 18.  
FT REPEAT 1261 1266 19.  
FT REPEAT 1270 1275 20.  
FT REPEAT 1279 1284 21.  
FT REPEAT 1288 1293 22.  
FT REPEAT 1297 1302 23.  
FT REPEAT 1416 1421 24.  
FT REPEAT 1425 1430 25.  
FT REPEAT 1434 1439 26.  
FT REPEAT 1443 1448 27.  
FT REPEAT 1452 1457 28.  
FT REPEAT 1461 1466 29.  
FT REPEAT 1470 1475 30.  
FT REPEAT 1479 1484 31.  
FT REPEAT 1488 1493 32.  
FT REPEAT 1497 1502 33.  
FT REPEAT 1616 1621 34.  
FT REPEAT 1625 1630 35.  
FT REPEAT 1634 1639 36.  
FT REPEAT 1643 1648 37.  
FT REPEAT 1652 1657 38.  
FT REPEAT 1661 1666 39.  
FT REPEAT 1670 1675 40.  
FT REPEAT 1679 1684 41.  
FT REPEAT 1688 1693 42.  
FT REPEAT 1697 1702 43.  
SO SEQUENCE 1829 AA; 197445 MW; 8F63506E1FED9840 CRC64;

Query Match 2.7%; Score 126.5; DB 1; Length 1829;  
Best Local Similarity 20.2%; Pred. No. 48;  
Matches 186; Conservative 86; Mismatches 306; Indels 345; Gaps 40;

QY 170 ARLEFDGYNKVRLEAKRFQSSGSETSAIGSSHOKTEPYA-----NIKALEDITQES 223  
DB 361 AREFFKGLPSFKDLAEKFRDLFPNPE-GWIDDGHOCLAPWVETKRNKGKHYHYD---PL 416

QY 224 AMDLNGSLPRLKOTALVARAVGYDDIDSLIRNSIGEY----VIHDLGPYYIIDRA 279  
DB 417 ALDLDDG---GIEFATVATGFGAGSLFDHTNNGIRATGWSADGLIVRDLNNGIJD--- 470  
QY 280 VEVRREGADKRFVTVADDEVPLLLIGDVFHHGKYEKKNIENASAEHGFQDGRMLDRSYD 339  
DB 471 ----NGAE-----LFED-----NTR--LADGSFAKHGTAALAELEDSKD 503  
QY 340 VILPNTAVSL-----IYDTGQYRDEVVEFTID----- 370  
DB 504 NII--NAADAAQTILRVWQDLDNQGDSQANELRTLEELIGQSIDLAYKQVKNKLGNGNTL 561  
QY 371 ----PKTNQLTTD-----PKLPYKRRELLQLTLVNNGEAVNQAQVA 409  
DB 562 AAGSSTYTKDGTAKKGDLLLAADNLHSREKKVELTAQAQAANLAGRLRLREANA 621  
QY 410 LSNDLIATRFYFMVMVTEIYFPEREQIONQVSFEQSSSRTPAQVDESLPEVETVEL 469  
DB 622 LSGDLA-----NML-----KASAAETKEAQL--ALDLNLHKMAE 655  
QY 470 TDGIILMDSPIEFSA-----SNLIQDKLNVA--KARHLYMPDDRVIA 512  
DB 656 TDSNNGKSPMLSTDWQTANEGIALTPSQVAQLKKNALVSLSDKAKAIDAARDRIAY 715  
QY 513 INHDDGVNRSILGRISDAVASAVARALPDESENEVIDLPERTALNARKTPADYQSKVP 572  
DB 716 LDAYTG-----QDSNTLYMSEEDALNIKYVNTDITDHLAKN 752  
QY 573 LY---VFVADKPRDQIGLGGSDT----- 595  
DB 753 IYONLLFQTRLQPYLNQISFKMENEDFTLDFSGLYQAFNHVKEINPQAKAFVLAEMLAYG 812  
QY 596 -----GTRLYT-----KEEHN-----LINRDGQAQAERL--SEBK 625  
DB 813 ELRSWYEGRRMLTDYVEEAKKAKGKEDYQKVLQETVALLARTSGIADDDILQNVGFGHN 872  
QY 626 KGVKLYATKP-----LSHPLNDOLRATLGYOEFYFHS-----TN 660  
DB 873 KNVSLGNDGNDTLIGAGNDYLEGSGSDTYFYGSGFQDYYNNDYATGRKDIIRFTD 932  
QY 661 GF--DLSTRPLE--HEISR-----SIIONGWNRYSLARYLRLDKLTQAPETW 705  
DB 933 GITADMLFTFBGNNHLLIKADGSGGVTVQSYFQNG-----SGAVIDE----- 977  
QY 706 QDLPVDFVNGKSGQELLAGVAVHKTVAADNLVPMMGYRQRYSLFEGS--SGLYSDANMA 763  
DB 978 ----IHFDMGKYLDAV-----TVKELVQOSTDGSRLYAYQSGNTLNGSLGD--- 1020  
QY 764 IARAGISGVYSFGDNAYGSNRAHOMTGGIOAGYIWSDNFNHVPYRLRFPAGDOSIRGYA 823  
DB 1021 -----DYLGXADGDLLNGDAGNDISYSGN-----GNDTLGGGEG 1055  
QY 824 HDLSLPSIDKGYITGGQVLAIVTAENYFEMKDLRLAVGDIQNAV--DKGFTNDTRI--GA 881  
DB 1056 NDALYVNGNDALNGGE---GNDHLNGEGDND---TLIGAGNDYLEGGSGSDTYVFGK 1108  
QY 882 GYGAVRWASPVGGVRVDYATGVKE 904  
DB 1109 GFGQDVT-----YNYDYATGRKD 1126

RESULT 37  
ID RPL\_HUMAN  
AC P56715;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Oxygen-regulated protein 1 (Retinitis pigmentosa Rpl protein)  
DE (Retinitis pigmentosa 1 protein).  
GN RPL OR ORPL.  
OS Homo sapiens (Human).



OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND VARIANTS HIS-872; TYR-985; THR-1670; PRO-1691  
RP AND TYR-2033.  
RC TISSUE=Retina;  
RX MEDLINE=99318096; PubMed=10391212;  
RA Sullivan L.S., Heckenlively J.R., Bowne S.J., Zuo J., Hide W.A.,  
RA Gal A., Denton M., Inglehearn C.F., Blanton S.H., Daiger S.P.;  
RT "Mutations in a novel retina-specific gene cause autosomal dominant  
RT retinitis pigmentosa";  
RL Nat. Genet. 22:255-259(1999).  
RN [2]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Retina;  
RX MEDLINE=99318095; PubMed=10391211;  
RA Pierce E.A., Quinn T., Meenan T., McGee T.L., Berson E.L., Dryja T.P.;  
RT "Mutations in a gene encoding a new oxygen-regulated photoreceptor  
RT protein cause dominant retinitis pigmentosa";  
RL Nat. Genet. 22:248-254(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99330563; PubMed=10401003;  
RA Gullionneau X., Piriev N.I., Danciger M., Kozak C.A., Cideciyan A.V.,  
RA Jacobson S.G., Farber D.B.;  
RT "A nonsense mutation in a novel gene is associated with retinitis  
RT pigmentosa in a family linked to the RPI locus";  
RL Hum. Mol. Genet. 8:1541-1546(1999).  
CC -1- FUNCTION: COULD HAVE A ROLE IN THE DIFFERENTIATION OF  
CC PHOTORECEPTOR CELLS.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN RETINA. NOT EXPRESSED IN HEART,  
CC BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, SPLEEN AND  
CC PANCREAS.  
CC -1- DISEASE: DEFECTS IN RPI CAUSE RETINITIS PIGMENTOSA FORM 1 (RPI); A  
CC DISEASE CHARACTERIZED BY CONSTRICTION OF THE VISUAL FIELDS, NIGHT  
CC BLINDNESS, AND FUNDUS CHANGES. THE DISEASE SEEMS TO BE ASSOCIATED  
CC WITH TRUNCATED (STOP OR FRAMESHIFT MUTATIONS) FORMS OF THE  
CC PROTEIN.  
CC -1- SIMILARITY: CONTAINS 2 DOUBLECORTIN DOMAINS.  
CC -1- DATABASE: NAME=Retnet;  
CC NAME=Retinal information network;  
CC WWW="http://www.sph.uth.tmc.edu/retnet/";  
CC -1- DATABASE: NAME=Mutations of the RPI gene;  
CC NOTE=Retina International's Scientific Newsletter;  
CC WWW="http://www.retina-international.com/sci-news/rpimut.htm".  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL: AF143226; AAD44197.1; -;  
DR EMBL: AF143224; AAD44197.1; JOINED.  
DR EMBL: AF143225; AAD44197.1; JOINED.  
DR EMBL: AF143223; AAD44198.1; -;  
DR EMBL: AF141021; AAD42072.1; -;  
DR EMBL: AF152242; AAD46774.1; -;  
DR EMBL: AF152240; AAD46774.1; JOINED.  
DR EMBL: AF152241; AAD46774.1; JOINED.  
DR EMBL: AF146592; AAD46769.1; -;  
DR Gene: HGNC:10263; RPI.  
DR MIM: 603937; -;  
DR MIM: 180100; -;  
DR InterPro: IPR003533; DCX.  
DR Pfam: PF03607; DCX; 2.  
DR SMART: SM00537; DCX; 2.  
DR PROSITE: PS03039; DC; 2.  
KW Vision; Retinitis pigmentosa; polymorphism; Repeat.  
FT DOMAIN 36 118 DOUBLECORTIN 1.  
FT

FT DOMAIN 154 233 DOUBLECORTIN 2.  
FT DOMAIN 268 273 POLY-SER.  
FT DOMAIN 671 675 POLY-LYS.  
FT DOMAIN 1687 1691 POLY-SER.  
FT VARIANT 872 872 R -> H.  
FT VARIANT 985 985 /FTid=VAR\_007810.  
FT VARIANT 1670 1670 N -> Y.  
FT VARIANT 1670 1670 /FTid=VAR\_007811.  
FT VARIANT 1691 1691 A -> T.  
FT VARIANT 1691 1691 /FTid=VAR\_007812.  
FT VARIANT 2033 2033 S -> P.  
FT VARIANT 2033 2033 /FTid=VAR\_007813.  
FT VARIANT 2033 2033 C -> Y.  
FT SEQUENCE 2156 AA; 240659 MW; 55ADBECA3D6A507 CRC64;  
SQ  
Query Match 2.7%; Score 126.5; DB 1; Length 2156;  
Best Local Similarity 17.5%; Pred. No. 62;  
Matches 181; Conservative 159; Mismatches 367; Indels 325; Gaps 42;  
QY 22 PLMTSQAQAQNNPANNITNHYPAHDTAINQAKAGNPVLLTPEDIQARLNAAGLNAPQOS 81  
DB 126 PMLSSRAISA-----HSPHPVAV-----AAGMPRPPRS 155  
QY 82 QALDVNFDQSPISRIQGEQSPPIGLQMSVTEETTPLSLELFAQESTEMGINS- 135  
DB 156 ----LVYFRNDDEPKTR-----RAVLISRRVYQSFAPLQHLTEVMQRPVVKLYA 200  
QY 136 --NDYIPEYQGEQNPSENVVPPTEPEKPGILKRLYARLFGDGVNK-----VPLKAKF 187  
DB 201 TDGRVPSLAQVILSSGAVNVAAGREPKG-----NYDIQKYLPLARLGISORV 250  
QY 188 YQSSQSGTSAIGSHQKTEPYANIKALLEDIQESADLNGSIPRLQALVA--ARAV 245  
DB 251 YPKGNASESKISTHSSSSRSQIYSSSEKTHNNCDLYSEVPEKYLALERNKDSQNL 310  
QY 246 GYV----DIDISTIRNSIGEVDYIIHDLGEVYIDYRAVEFRGAGDAKAF--TYVA-- 296  
DB 311 PIYSEDDIEKSTIFPDGDTWT-----EMKRFRIKEETIKMTTYSKT 356  
QY 297 ----DEVPLIGDVFHNGKYEYTK-----NLINASAEHGYFDGRMLDRS 337  
DB 357 GPSNNDK-----SEMSPPGRTESSSGKLACFSADVSMEMSSNDQGS-----LAEE 407  
QY 338 VDVILPNTADVSLIYDTGYQRFDEVVFTIDRKTQNLTPDPKLPVKRELLBOLLTVN 397  
DB 408 INIQMTQVA-----ETCSASWENA---TVPDIIQGTQDQAKHFRYPRTGLRVR 458  
QY 398 MGEVYNQAVRALSNDLIATRYFMVTELYFPEREIQNDQV-----SFEQSSSRTE 451  
DB 459 QKRSV-IGSVTLVSEFEVQK--MIQFYSSEEREGENKSEYHMFTHSCSKMSVSNK 514  
QY 452 PA-----QVDESTLEPTE-----TYELTGILMDIS-----PIEFASN 486  
DB 515 PLVVOINNNDQMESSSLERKKNSILKSSAISAGVITITQKMLKMSHNNGLDSTISNNS 574  
QY 487 LIDDKL-----NLVAAKARHLVDPDDRVLAINH-----DGVNRSI-----LGRI 527  
DB 575 IVEEDVVDVLDNKTGIGIKFKFTYGTNDRFSLADATHFSSNNGSTDKNISAPARSEA 634  
QY 528 SDVAVSAVARAILPDESENEVIDLPKRALANRKTTPADVYOSKVPVLYVYASOKPRDQI 587  
DB 635 SSTVTARIDRLINEFAQCGLTKLPKN---EKKILSSVASKRKRKSQAQINSRYODGOL 690  
QY 588 GLQWGSDDTGRVLTGK--FEHNILNRD-----GYQAGAEPLRSE----- 623  
DB 691 ATKGIINKMERINTKGRITKEMATVODSDSLKGCILCEBDLQASDVYIESNTGCSNSLN 750  
QY 624 -----DKGVKLYATKPLSHPLND-----QLRATLGYQOEVFGHSTNGFD 663  
DB 751 STISKNFHNRKLNLTQNSKVGGLLTRKRSRLNKISLGAAPKKEIQGRKRVFPHN----- 805  
QY 664 LSTRILEHLSIRSIIONGGMNRTYSLRRLDKLKTQAPETWODLPVDVYVNGKPSQDALL 723

```
Db 806 -----ESKCYKSTFEENKSLFHVFNILEQ--KPKDFYAPQSOAEVAGSYLRG-----M 850
Oy 724 AGVAVHKTVANLVNPMGYRORYSLEVGSSGLVS-----DANMAIAR 766
Db 851 AKKSILVSKVTOSSHIT-LKSQKKRKKDKYKASAILSKOHATTRANSLSLKRPFEALIH 909
Oy 767 AGI-----SGVYSFGDNAYGNSNRHOMTGTGIGIAGYI 797
Db 910 HSIQNYIGISWLOININPYPTLPKIPKSAFYCRNETSVYNCSSNFSGNDPHTNSGKI-SNFFV 968
Oy 798 WSDNHNHPRYLRFPAAGDQSRGVAHDSLPSIDKGYLTGGQVLAVGTAEVNYEFMKDL 857
Db 969 MESN-KHI-----TKIAGLTGDNLCKEGDKSF-----IANDTGE-----EDL 1004
Oy 858 RLAVFGDIGNAY 869
Db 1005 HETGVGSLNDAY 1016

RESULT 38
YLJ2_CAEEL STANDARD; PRT: 2198 AA.
ID YLJ2_CAEEL STANDARD; PRT: 2198 AA.
AC P34367;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 256.3 kDa protein C50C3.2 in chromosome III.
GN C50C3.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RA [1]
RA SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardiner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Lareille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smailson N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RA elegans."
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: L14433; AAA27973.1;
CC WormPep: C50C3.2; CE01861.
CC InterPro: IPR002048; EF-hand.
CC InterPro: IPR002017; Spectrin.
CC Pfam: PF00036; EFhand; 2.
CC Pfam: PF00435; Spectrin; 10.
CC ProDom: PD000012; EF-hand; 1.
CC SMART: SM00150; SPEC; 5.
CC PROSITE: PS00018; EF_HAND; 1.
KW Hypothetical protein, Calcium-binding; Repeat.
FT CA_BIND 2025 2036 EF_HAND (POTENTIAL).
SO SEQUENCE 2198 AA; 256260 MW; CDB2C1092C5EDC8 CRC64;
```

```
Query Match 2.7%; Score 126.5; DB 1; Length 2198;
Best local similarity 18.7%; Pred. No. 64;
Matches 124; Conservative 101; Mismatches 248; Indels 189; Gaps 27;

Oy 81 SGALDVNFDQSPISR--IGGOSPPGLGDMVIEETPLSLLELFAOESTMGINPDY 138
Db 1232 NDSMSQILIDEPDITRQIKETPALEADVSNKTRPDIERM-QKRHDELAJDINDY 1290
Oy 139 IPEYGEQPNSEVVPPTLEPEKPGILKRLYARLFNDGVNKPRLKARFYQSSQ-----S 193
Db 1291 HTIYEDFLNNGVY-----EEEKVQVIRHLWSSLIESTTTPKPSI-AQCQKSRFLFELD 1343
Oy 194 GFTSAIGSSHQKTEPYANIKALIEDITQESAMDLDNGSIPRL--KQTAIVARAAGYDI 250
Db 1344 GMSIWLIDAETDVHSTTFEIGLYNSVDAEKARLLVSLGEQAGEKDDILIRIQCESVDY 1403
Oy 251 DLSIRNSIGEVDYIIHIDGEPVYIDYRAVEYRGSGADKAPFTVADEVPLL--IGDVFH 308
Db 1404 IYERLRNGISQLOMLIKANNEDLV-WKSMQ-----KYVTAIDDELCEWKEINVIFS 1454
Oy 309 HGKY--ETRKNLIENASAEHGYFDGRWLDKRSVDVILPDNTADVSLIYDTGYRDEYVF 366
Db 1455 SYNVGNDVSSNDVLRKQHRLQLEFORQOKYAKVYLLTTELVSRRPSLEYKFEV-- 1511
Oy 367 FTIDPKTQNTDTPDKLPYKRELLFOLLVNMGEAVNIGQAVANALNDLIA--TRYENMVT 425
Db 1512 --IDKRVAVLT-----ELLSS-----NRQIAIRIKRLKEKTEYVTNME 1549
Oy 426 EIVPFEREQIQNDQVSEFQSSSRTEPAQVDESTLEPVLETVDGLMDISPIEFSAS 485
Db 1550 ---IREKEGLL-DQLLDKLTGPEYVLADVERKL--QVLETV----- 1585
Oy 486 NLIQKLNIVAAKARHLMDPDDRYLAINHDGVNRSILGRISDAVSANARAILDESEN 545
Db 1586 ---GDHMDLTITKAE---QMSDDEVY-----RTKVAVTAI-----DQLRK 1619
Oy 546 EYIDLPERALNRKTPADVQSKKVPYLFVYASDKPRDQIGLGMSGSTGRLYTKFEB 605
Db 1620 KVLKNNELKKNHOKIKESIETK-----FVST-----CDTGIOCI----- 1655
Oy 606 NLINDGYQAGAEELRLSEDKKGVKLYATKPLSLPNDQIRATL---GYQGVF----- 655
Db 1656 -----REQERKIGNLVHQRPAENFENTAYHSTMTVEYTKDTEFLKVY 1701
Oy 656 -----GHSTNG-----FPLSTRTEHEHSRTIQNG 681
Db 1702 CKRLQNSTDAEKRLSLVSERLNALKKQDLDLAEKIAVDKFKLVQNIODEYSRTACEIG 1761
Oy 682 GW 683
Db 1762 NW 1763

RESULT 39
PCGV_HUMAN STANDARD; PRT: 3396 AA.
ID PCGV_HUMAN STANDARD; PRT: 3396 AA.
AC P13611; P20754; Q9UNW5; Q13010; Q13189; Q15123;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Versican core protein precursor (large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glia)
DE hyaluronate-binding protein) (GHAP).
GN CSB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM VO).
RX MEDLINE=95105188; PubMed=7528742;
RA Naso M.F., Zimmermann D.R., Iozzo R.V.;
```

RT "Characterization of the complete genomic structure of the human  
 RL versican gene and functional analysis of its promoter.";  
 RN J. Biol. Chem. 269:32999-33008(1994).  
 [2]  
 RP SEQUENCE FROM N.A. (ISOFORM V1).  
 RC TISSUE-Placenta;  
 RX MEDLINE=90059882; PubMed=2583089;  
 RA Zimmermann D.R., Ruoslahti E.;  
 RT "Multiple domains of the large fibroblast proteoglycan, versican.";  
 RL EMBO J. 8:2975-2981(1989).  
 [3]  
 RP SEQUENCE FROM N.A. (ISOFORM V2).  
 RC TISSUE-Glia; tumor;  
 RX MEDLINE=95105187; PubMed=7806529;  
 RA Dours-Zimmermann M.T., Zimmermann D.R.;  
 RT "A novel glycosaminoglycan attachment domain identified in two  
 RL alternative splice variants of human versican.";  
 RN J. Biol. Chem. 269:32992-32998(1994).  
 [4]  
 RP SEQUENCE OF 2711-3396 FROM N.A.  
 RC TISSUE-Lung fibroblast;  
 RX MEDLINE=88007514; PubMed=2820964;  
 RA Krusius T., Gehlsen K.R., Ruoslahti E.;  
 RT "A fibroblast chondroitin sulfate proteoglycan core protein contains  
 RL lectin-like and growth factor-like sequences.";  
 RN J. Biol. Chem. 262:13120-13125(1987).  
 [5]  
 RP SEQUENCE OF 251-347 FROM N.A.  
 RX MEDLINE=93122792; PubMed=1478664;  
 RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,  
 RT McPherson J.D.;  
 RL "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of  
 RN human chromosome 5 (5q12-5q14).";  
 [6]  
 RP SEQUENCE FROM N.A. (ISOFORM V3).  
 RC TISSUE-Brain;  
 RX MEDLINE=95181355; PubMed=7876137;  
 RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;  
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M  
 RL without a chondroitin sulfate attachment in region in mouse and human  
 RN tissues.";  
 RL J. Biol. Chem. 270:3914-3918(1995).  
 [7]  
 RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).  
 RC TISSUE-Aortic smooth muscle;  
 RX MEDLINE=99327053; PubMed=10397680;  
 RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,  
 RT "Versican/PG-M isoforms in vascular smooth muscle cells.";  
 RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).  
 [8]  
 RP PARTIAL SEQUENCE.  
 RC TISSUE-Brain;  
 RX MEDLINE=89174663; PubMed=2466833;  
 RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;  
 RT "Isolation and partial characterization of a glial  
 RL hyaluronate-binding protein.";  
 RN J. Biol. Chem. 264:5981-5987(1989).  
 [9]  
 RP TISSUE SPECIFICITY OF ISOFORMS.  
 RX MEDLINE=96213482; PubMed=8627343;  
 RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;  
 RT "Differential expression of versican isoforms in brain tumors.";  
 RL J. Neurooncol. Exp. Neurol. 55:528-533(1996).  
 CC -1- FUNCTION: May play a role in intercellular signaling and in  
 CC connecting cells with the extracellular matrix. May take part in  
 CC the regulation of cell motility, growth and differentiation. Binds  
 CC hyaluronic acid.  
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
 CC -1- ALTERNATIVE PRODUCTS: At least 5 isoforms: V0 (shown here), V1,  
 CC V2, V3 and VINT; are produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed

CC in normal brain, gliomas, medulloblastomas, schwannomas,  
 CC neurofibromas, and meningiomas; v2 is restricted to normal brain  
 CC and gliomas; v3 is found in all these tissues except  
 CC medulloblastomas.  
 CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development.  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U16306; AAA65018.1; -;  
 CC EMBL: X15998; CAA34128.1; -;  
 CC EMBL: S52488; AAB24878.1; -;  
 CC EMBL: U26555; AAA67565.1; -;  
 CC EMBL: D32039; BAA06801.1; -;  
 CC EMBL: J02814; AAA36437.1; -;  
 CC EMBL: AF084545; AAD48545.1; -;  
 CC PIR: S06014; S06014.  
 CC PIR: A29348; A29348.  
 CC PIR: A30358; A30358.  
 CC HSSP: P01132; 1EGF.  
 CC Genew: HGNC:2464; CSPG2.  
 CC MIM: 118661; -;  
 CC InterPro: IPR000152; Asx\_hydroxyl.  
 CC InterPro: IPR000561; EGF-like.  
 CC InterPro: IPR000742; EGF-2.  
 CC InterPro: IPR001881; EGF\_CA.  
 CC InterPro: IPR001438; EGF\_TI.  
 CC InterPro: IPR003599; Ig.  
 CC InterPro: IPR003006; Ig\_MHC.  
 CC InterPro: IPR001304; Lectin\_C.  
 CC InterPro: IPR000538; Link.  
 CC InterPro: IPR000436; Sushl\_SCR\_CCP.  
 CC Pfam: PF00008; EGF\_2.  
 CC Pfam: PF00047; Ig\_1.  
 CC Pfam: PF00059; lectin\_C\_1.  
 CC Pfam: PF00084; sushl\_1.  
 CC Pfam: PF00193; Xlink\_2.  
 CC PRINTS: PR00010; EGFBL00D.  
 CC ProDom: PD000918; Link\_2.  
 CC SMART: SM00032; CCP\_1.  
 CC SMART: SM00034; CLECT\_1.  
 CC SMART: SM00179; EGF\_CA\_1.  
 CC SMART: SM00001; EGF\_like\_1.  
 CC SMART: SM00409; IG\_1.  
 CC SMART: SM00445; LINK\_2.  
 CC PROSITE: PS00010; ASX\_HYDROXYL\_1.  
 CC PROSITE: PS00022; EGF\_1\_2.  
 CC PROSITE: PS0186; EGF\_2\_1.  
 CC PROSITE: PS01187; EGF\_CA\_1.  
 CC PROSITE: PS01241; LINK\_2.  
 CC PROSITE: PS00615; C\_TYPE\_LLECTIN\_1\_1.  
 CC PROSITE: PS50041; C\_TYPE\_LLECTIN\_2\_1.  
 CC Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushl;  
 CC Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
 CC Hyaluronic acid; Alternative splicing.  
 CC FT CHAIN 1 3396 POTENTIAL.  
 CC FT DOMAIN 37 137 VERSICAN CORE PROTEIN.  
 CC FT DOMAIN 167 244 IG-LIKE V-TYPE DOMAIN.  
 CC FT DOMAIN 265 346 LINK 1.  
 CC FT DOMAIN 348 1335 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT  
 CC DOMAIN).

```

FT DOMAIN 1336 3089 GAG-BETA.
FT DOMAIN 3089 3125 EGF-LIKE 1.
FT DOMAIN 3125 3163 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3163 3290 C-TYPE LECTIN.
FT DOMAIN 3290 3353 SUSHI.
FT DISULFID 44 130 BY SIMILARITY.
FT DISULFID 172 243 BY SIMILARITY.
FT DISULFID 196 217 BY SIMILARITY.
FT DISULFID 270 345 BY SIMILARITY.
FT DISULFID 294 315 BY SIMILARITY.
FT DISULFID 3093 3104 BY SIMILARITY.
FT DISULFID 3098 3113 BY SIMILARITY.
FT DISULFID 3115 3124 BY SIMILARITY.
FT DISULFID 3131 3142 BY SIMILARITY.
FT DISULFID 3136 3151 BY SIMILARITY.
FT DISULFID 3153 3162 BY SIMILARITY.
FT DISULFID 3169 3180 BY SIMILARITY.
FT DISULFID 3197 3289 BY SIMILARITY.
FT DISULFID 3265 3281 BY SIMILARITY.
FT DISULFID 3296 3339 BY SIMILARITY.
FT DISULFID 3325 3352 BY SIMILARITY.
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 809 809 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1332 1332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1338 1338 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1442 1442 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1468 1468 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1663 1663 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1898 1898 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2179 2179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2272 2272 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2280 2280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2360 2360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2385 2385 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 2.74; Score 126.5; DB 1; Length 3396;
Best Local Similarity 19.8%; Pred. No. 1.2e+02;
Matches 153; Conservative 118; Mismatches 293; Indels 207; Gaps 39;

66 IOARINAGLAKPO---SQALDVNFPDPSIRIGOSPGLGDMVIEET--PLSL 120
358 ILAEFASLSLSEKPMQWMDRTPIIPLVDLPLV--IPPEFPVGNIVSEKATVQPOAI 415
121 EELFAOE-----STEMGINPDYIPYOGEOPNSEVVVPTLEPEKGLRLYARLFN 174
416 TDSLATKLPPTGTSTYKPKWMDMDVSPSASGPL----- 447
175 DGVNKPPLKAKFYOSOGESATIGSSHOKPEPYANI---KALEDITQ-----E 222
448 -GKLDISEKKEEVLDT-----TGVSHYATDSWDGVYEDKQTESYQIQIEVGPLV 499
223 SAMDLNGSIPR---LRQATVAARAVGYIDLSIRNSIGEVDYI--DLGEPYVI 275
500 TSMELKHPKSPKPEFVTEPLVTARMLLESKTEKMW--STVELVTTGHYGTLGEBDE 558
276 DYRAVEVGEAGDKAFKFTTVADEVPLI-----GDVPHNGKTEYKKNLENAASHGTF 329
559 D-RTLTV--GSDS--LFIPOIEVITVSKTSEDTIH-----THLEDELSVAS-- 603
330 DGRWLDKRSVDVILPNT--ADVSLIVDTGTOYRFDEVV--FFTIDPKTQTLTDDKLPV 385
604 -----TYSPLIMPONNGSSMDMWEROTSGRTTEEPFGKYSTTRPPSQHTEIELFPY 658
386 KRE--LLEQLTV-----NGEAYNLQAVRALSNDLIATRYNNMVTETV--FPERQION 437
659 SGDKLIVGIGIVTYPSTLOTETTHREERTETLIPKARDTYVDEIOELETKSPFMKTEE 718
438 DQVSEFSSSSSTPEPAOVDESFLPEVETVELTDLGILMDISPIESASNLIDKLNLVA 497
719 EVFSGMKLSTLSLEPIHYTES-----SVEMTK-----SDEPITLI---TKLSA 758

```

```

QY 498 KARHLYMPDDRVL---AINHDDGVNNSIL--GRISDAVSAVARAILPDESENEVIDLPE 552
DB 759 EPEVRMEEDFATPTGTCTDENITTVLLAHGTLSEAAVSKWSW--DEDNNTSKPLES 817
QY 553 RTALANRKPADVY-----QSKKVPVYFVASDK-----PRDQOIGLGMGSDTGTRLVT 601
DB 818 TEPASASKLPALLTGYGMKNCKDKDIPSFTEGDADDEFILIDSQOKOLEVTDIDIAHG 877
QY 602 KF--EHNLDNDGYQAGAEIRLSEDKGV-----KLVTAKPLSHPLNDQIATATGQ 651
DB 878 KFTIRFOPTSTGIAEKSLTLDSTTEKVPITSTEGQVATMGSS-----ALGEV 928
QY 652 QEV-----FGHSTN-----GFDLSTRTLE-----HEFSRSIIONGWNRTYS 688
DB 929 EDVLDKRPVSTVPOFAHTEVEGALFVYSSTOEFTTVYDSSHITPLSVIKRTDMG----- 984
QY 689 LRVRLDKLTQAPPETWODLPVDFVNGKPSOBALLAGVAHVKTADNLVNP 739
DB 985 -----VLVPSVPSF-----DEVLGERSQDILV-----IDOTRLREATISP 1018

```

```

RESULT 40
ID P200_MYCGE STANDARD; PRT: 1616 AA.
AC 049429; 049259; 049298; 049352; 049353;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein P200.
GN M3386.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A. / G-37;
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fitchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier J.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RA "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 256-427; 432-543 AND 1083-1140 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;
RA "A survey of the Mycoplasma genitalium genome by using random
RA sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -I- FUNCTION: COULD BE A ACCESSORY STRUCTURAL COMPONENT IN
CC CYTADHERENCE (BY SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
CC EMBL: U39720; AAC71613.1; -
CC EMBL: U02245; AAA03401.1; ALT_FRAME.
CC EMBL: U02175; AAD12458.1; -
CC EMBL: U02126; AAD12402.1; -
CC TIGR: MG386; -
CC Cytadherence; Structural protein; Repeat; Complete proteome.
FT DOMAIN 1205 1389 2 X 32 AA REPEAT.

```

```

FT REPEAT 1205 1236 1-1.
FT REPEAT 1358 1389 1-2.
FT DOMAIN 891 1389 2 X 26 AA REPEAT.
FT REPEAT 1161 1186 2-1.
FT REPEAT 1310 1339 2-2.
FT CONFLICT 256 256 P -> S (IN REF. 2).
FT CONFLICT 304 304 S -> F (IN REF. 2).
SQ SEQUENCE 1616 AA: 185678 MW: 6AF76A13AC49E4FF CRC64;

```

## Query Match

2.7% Score 125.5; DB 1; Length 1616;

Best Local Similarity 19.5%; Pred. No. 46; Mismatches 180; Indels 183; Gaps 31;

Matches 111; Conservative 94; Mismatches 180; Indels 183; Gaps 31;

```

QY 76 NAKPQAGLDVNFDDSPISRGQSPPLGLDMSVIEETPLSLLELFAQESTMG--- 132
DB 297 NTKSLASSL-----ENLLLENSSEQP-----VIEVAPRRNEVIFRNPVTKLHEK 344
QY 133 -----INP-----NDYIP--EYQEQ--PNSVVPVPTLEPE---KPLG---IKRLYARL 172
DB 345 EKEEFLNPVKETMETIPLIEIVKEEVESEVEAPVEIEPEACEPETIPEVETVF--V 402
QY 173 FNDGV-----NKVPRLKAF-YQSSQSEFSAIGSSHOKTEPYANIKALEDITQ 221
DB 403 YEDDLKGLDGNQOAGVPEVEYFVEDLKGLDSTIKDDQOHDE---IAKHVELLSQ 458
QY 222 ESAMDINGSIPRLQGTALVARAVGYDIDLSTIRNSIGEV-----DVIIDHIGEPV 273
DB 459 DYKKEIKDS-----AKA-----DLSNISDDIDSVWKEFGSFTDETOKSVEKS 501
QY 274 YIYIRAVEVGEQADKAF--TTVADEVPLLIGVFHHGKYET--KKNLIENASAEHGFD 330
DB 502 QVDEIILDANNDEINELFRDEVVNNIDSOINETVSEQEFPYVNEFQEFSEPVVD 561
QY 331 GRW-----LDRSV-----DVILPDNTADVSLIYDGTQYR-----FD 362
DB 562 EKIKETNSDESVTDLTALFSEKLVNEVLLTNEYVDVNAAPPSTETEVKVSSELPKSELVD 621
QY 363 EVVFTIDPKTNQ-----LFTDPDKL-----P-VKRELLEQLLTVN-MG 399
DB 622 EITFFINNDPQPGLEKVDLFLETEPKSLPDEKTTIVSESEPPFIQPDLSLELDVNDVD 681
QY 400 EAVNLQAVRA-LSNDLIATRYFNM-----VNTETVPEREQIONQVSF 442
DB 682 KSLKETTSVELNHEEIGNEFINLDVSEKEVQOQPTQLETDSFVLPYQIVED---SF 738
QY 443 EOSSSRTEPAQVDESTLEPVIEFVELTDGILMDISPIESASNL-----IQDK 491
DB 739 TESAEIPNEFSSEQKDTLEFISQTOE-----VETSESNAVPTVEQETKLFERHODE 787
QY 492 LNLVA-----AKARHLYD-MPDDR 509
DB 788 NNLFTPLPLDLFEIIESNMLFDSKPEDK 815

```

Search completed: April 28, 2003, 16:22:31  
Job time : 67 secs



GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:16:49 ; Search time 45 Seconds  
(without alignments)  
4207.940 Million cell updates/sec

Title: US-09-914-168-2

Perfect score: 4727

Sequence: 1 MSKPVLFANRSEMPALAAV.....TGYKEGNPIKLFHFGTTPF 919

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP viirus:\*  
16: SP bacteriap:\*  
17: SP archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	615.5	13.0	579	16	Q910U1 pseudomonas
2	532	11.3	586	16	Q9CK26 pasteurrella
3	494	10.3	578	16	Q82B95 yersinia pe
4	479	10.1	577	16	Q8XEL5 salmonella
5	454	9.6	582	16	Q9KP31 vibrio chol
6	353.5	7.5	593	2	Q9EY32
7	347	7.3	617	16	Q9PDZ7 xanthomonas
8	262.5	5.6	678	16	Q8RHT4 xylella fas
9	259.5	5.5	792	2	Q95359 fusobacteri
10	259	5.5	797	16	Q9K1H0 neisseria g
11	259	5.5	797	16	Q9JX31 neisseria m
12	254	5.4	798	2	Q8VQ24 bartonella
13	253	5.4	797	2	Q30912 neisseria m
14	244.5	5.2	765	16	Q8XZ13 raietonia s
15	242.5	5.1	589	16	Q8XWT6 raietonia s
16	236.5	5.0	641	16	Q8UC83 agrobacteri

17	232.5	4.9	774	16	Q8UF16 agrobacteri
18	229.5	4.9	793	2	Q93PM2 haemophilus
19	225	4.8	861	16	P73472 synechocyst
20	223.5	4.7	781	16	Q8YH0 bruceella me
21	223.5	4.7	794	16	Q98MC3 rhizobium 1
22	220	4.7	615	16	Q9JWM5 neisseria m
23	220	4.7	617	16	Q98K30 rhizobium 1
24	220	4.7	618	16	Q92LN9 rhizobium m
25	216.5	4.6	739	16	Q9PIZ8 campylobact
26	216	4.6	628	16	Q9ATW6 cauliobacter
27	215	4.5	635	16	Q9JXB7 neisseria m
28	214	4.5	784	16	Q9PE12 xylella fas
29	211.5	4.5	795	16	Q8ZHS8 yersinia pe
30	209	4.4	623	16	Q8YET3 bruceella me
31	206.5	4.4	916	16	Q25369 helicobacte
32	205.5	4.3	768	16	Q9ZEB3 rickettsia
33	202.5	4.3	768	16	Q92J67 rickettsia
34	198	4.2	738	2	Q9X3V8 zymomonas m
35	197.5	4.2	833	16	Q8YUR6 anabaena sp
36	197	4.2	776	16	Q92Q48 rhizobium m
37	194.5	4.1	906	16	Q9ZLH9 anabaena sp
38	194	4.1	789	2	Q51930 pasteurella
39	193.5	4.1	676	16	Q8YWP0 anabaena sp
40	190.5	4.0	769	16	Q9A711 cauliobacter
41	183.5	3.9	797	16	Q9HXY4 pseudomonas
42	181.5	3.8	797	2	Q9S341 photorhabdu
43	180	3.8	782	2	Q44629 bruceella ab
44	177.5	3.8	778	16	Q67326 quifex aeo
45	177.5	3.8	803	16	Q9KFW0 vibrio chol

## ALIGNMENTS

## RESULT 1

ID Q910U1 PRELIMINARY; PRT; 579 AA.

AC Q910U1;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)

DE Hypothetical protein PA2543.

GN PA2543.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI\_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC SRRIN-ATCC 15692 / PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

opportunistic pathogen.";

RL Nature 406:959-964 (2000).

DR EMBL: AF004682; AAC05931.1; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 579 AA; 63739 MM; 23467005C836FBBF CRC64;

Query Match 13.0%; Score 615.5; DB 16; Length 579;

Best Local Similarity 24.5%; Pred. No. 4.3e-28;

Matches 178; Conservative 117; Mismatches 220; Indels 213; Gaps 17;

QY 210 ANIKAALEDITQESAMDLNGSIPRLQRTALV---AARAVGYDIDLSITRNSIGEDVYI 265

DB 41 ANIEAVYVSLGERD---EALQGRFRNMAEQAEKAAQALGYRQ-----AQIDSE 86

QY 266 IHD-----LGEPIYIDIRAVEVGECAADKAFITVADEPVLIDVFRHCKYET 314

```
Db      87 VKDGKPKLTLLKVVYGEFRLROVNIQVLGEAASLESFRLPSGK-OLKPGAKLNGVED 145
Qy      315 KKNLENSAEHGYFDGMDRSVDVILPDMNTADSLYDGTGYRPFPEVFTIDPKTN 374
Db      146 AKRLIQONASRGFFQGRSTORLSDIPRAGIADIDLVDYSQRTTFGKVSF----- 197
Qy      375 QLTDPDKLPVKRELLEQLLTVMNGEAYNLQAVRALSNDLIATRYFNNVNTETIVEPEREQ 434
Db      198 ----DGDST-IEEELLRRVVPFKAGQPYDSIELIENQNLQSGVF----- 238
Qy      435 IONDOVSEEOSSSRTPEAOVDESTLEPIETVELTDGLMDISPIESASNLIDKLN 494
Db      239 -----EGVRDAAPTQQA----- 252
Qy      495 VAAKARHLVMPDDRVLAINHDDGVNRSILGRISDAVASAVARAILPDESENEVIDLPERT 554
Db      253 -----DGAQA----- 259
Qy      555 ALANKRTPADVYQSKKVPVLYFVASDKPRDQIGLGWGSDTGTRLVTKFEHNLINRDGYO 614
Db      260 -----PVAVRLERAKPRMTGVLGFSTDVGARARFNTRRHVAEGHS 302
Qy      615 AGAELRLSEDKKGVKLYATKPLSHPLNDQLRATLGQOEVEFGHSTNGPDLSTRTI---- 670
Db      303 LGFESEISAPRONVGAMYEIPLDPLTLDKLRFTSGTQFE-----DLVDTESKLTLTGGE 356
Qy      671 HEISRSIIONGGMNRTYSLRYRLDKLTQAPPETWQDLPVDFVNGKPS--QOALLAGVAV 728
Db      357 WHSKRP-----DCMQRVVSLMMREYKLT-----GDDSGLSFLMPCIGY 396
Qy      729 KHTVADNLVNPFRGYRQKRSLEEVSSGLVSDANMALIARAGISGVYFSGNATVGSRAHOM 788
Db      397 SLEETDNKNVDPHSGYRLQFNVKGAREGLADADVLHVDMAGLTSFAG-----GHRLL 449
Qy      789 TGGIAGYIYSNFMHVPYRLRFEPAGDQISIRGYAHDLSPISDGYTLGGVLAVGTA 848
Db      450 IGRLOYGATATDYKSTPSLRFEPAGDQSVRGYRTLSPEMSODKIGGKRYMAGSV 509
Qy      849 YNFERMKDLRLAVFGDIGNAYDKFTNDTKIGAGVGYRMASPVGOVRVDVATGVKEEGNP 908
Db      510 YQYPLAERWRRLATFVDQGNAFNSLDPSIKTGVGFGVWVSPVGLRLDLALALDDDG- 568
Qy      909 IKLHFEIG 916
Db      569 FRLHFSMG 576

RESULT 2
O9CK26 PRELIMINARY: PRT: 586 AA.
AC O9CK26:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein PM1809.
GN PM1809.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PM70;
MEDLINE=21145866; PubMed=11248100;
RA May B.-J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL: AE006218: AAK03893.1: -
GN Interpro: IPR001084: Bac_surfa9_D15.
DR Interpro: IPR001092: H4H_basic.
DR Pfam: PF01103: Bac_surfa9_Ag_1.
DR PROSITE: PS00038: HELIX_LOOP_HELIX: UNKNOWN_1.
```

```
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 586 AA; 67106 MW; 00895174DED58283 CRC64;
Query Match 11.3%; Score 532; DB 16; Length 586;
Best Local Similarity 23.0%; Pred. No.3.8e-23;
Matches 166; Conservative 107; Mismatches 258; Indels 192; Gaps 14;

Qy 204 OKTEPYANIKAALEDITQESANDLNGSIPRLKOTALVAARAVGY--DIDLSIRNSIGE 261
Db 41 QNTQNNNVRIYVGMIDKEADGSEHKKQLYREALDKLARAGYQSEVEFQIESQKPK 100
Qy 262 VD-VIIH-DLGEPPYIDYRAVEYRGEADDKAFTTYVADEPRLIDGVPHGKYEETKNNI 319
Db 101 KQQLIAHAKYIGKPKIADYDEVYLGQATODPEFIATVKIIPK-GSILNHEYYDNKSHI 159
Qy 320 ENASAEHGYFDGRWLDKRSVDVILPDMNTADSLYDGTGYRPFPEVFTIDPKTNQLTDT 379
Db 160 QKIALSRGYFDGDFLYSLREVRPSTQQAAMRLDYDSGGEYRFGGEYTF----- 206
Qy 380 PDKLPVKRELLEQLLTVMNGEAYNLQAVRALSNDLIATRYFNNVNTETIVEPEREQINDQ 439
Db 207 -ENAOIREDYELNNMIFNKGQPYLINDSTLTNNYSSNWFSSVLMQPYLDEEHKIVND 265
Qy 440 VFEEOSSSRTPEAOVDESTLEPIETVELTDGLMDISPIESASNLIDKLNVAAGA 499
Db 266 V-----LLQPR----- 271
Qy 500 RHLVMPDDRVLAINHDDGVNRSILGRISDAVASAVARAILPDESENEVIDLPERTALANR 559
Db 272 ----- 271
Qy 560 KTRPADVYQSKKVPVLYFVASDKPRDQIGLGWGSDTGTRLVTKFEHNLINRDGYOAGEL 619
Db 272 -----KKNSMEVGIWASDGPRLQGLGKTPWLNNGHSFRNL 310
Qy 620 RLSEDKKGVK-LYATKPLSHPLNDQLRATLGQOEVEFGHSTNGFDLS--TRTLEHISR 675
Db 311 YVSARQQLTEATYKMPMLKNPLNTYYEYSAGLENE-NKNDTSPFASLSAIAYWNNE--- 366
Qy 676 STIONGMNRTYSLRYRLDKLTQAPPETWQDLPVDFVNGKPSQOELLA--GVAHKTIVA 733
Db 367 -----AGWQHSLSGLRYRYD-----SFIDANVKDKTLLEPTASVARTRL 405
Qy 734 DNLVNPFRGYRQKRSLEEVSSGLVSDANMALIARAGISGVYFSGNATVGSRAHOMGTIO 793
Db 406 QCGLEPTTWGDTOKLTITDLRTWMLSDVDLKMGGSSLMVRYTLQH-----HRTYRL 458
Qy 794 AGYIWSDNFNHVPYRLRFEPAGDQISIRGYAHDLSPISDGYTLGGVLAVGTAEYNEF 853
Db 459 LGMLHTKNIERILPPALRFEPAGDQISIRGYKRIAPKNNAKLVGSRLLTGSFEYOYOV 518
Qy 854 MKDLRLAVFGDIGNAYDKFTNDTKIGAGVGYRMASPVGOVRVDVATGVKEEGNP IKLHF 913
Db 519 YPDWMLATPADTGLANQFTTKELRYGAGMGVWASPVGAIKFIDATFVDRDKNSKIOF 578
Qy 914 FIG 916
Db 579 YTG 581

RESULT 3
O8ZB95 PRELIMINARY: PRT: 578 AA.
AC O8ZB95:
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative exported protein.
GN YP03524.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
```



RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CO-92 / BIOVAR ORIENTALIS;  
 RA MEDLINE-21470413; PubMed-11586360;  
 RA Parhill J., Wren B.W., Thomson N.R.,  
 RA Prentice M.B., Sebaita M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Felwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;  
 RA "Genome sequence of Yersinia pestis, the causative agent of plague,"  
 RL Nature 413:523-527(2001).  
 DR EMBL: AJ414157; CAC92753.1; -  
 DR InterPro: IPR000566; Lipooln\_cytfABP.  
 DR PROSITE: PS00213; LIPOCALIN: UNKNOWN\_1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 578 AA; 64405 MW; 69AD3780BFC20F1 CRC64;

Query Match 10.5%; Score 494; DB 16; Length 578;  
 Best Local Similarity 23.4%; Pred. No. 6.5e-21;  
 Matches 167; Conservative 105; Mismatches 255; Indels 188; Gaps 17;

QY 211 NTKALEDI-TOSAMDLNLSI-PRLRQTALVAAKAVGYD--IDLSTINSGEVDYII 266  
 DB 38 NVARALSTIGTDEVTD--GRFRSRVDEALRQGLRALGYDPTTFELQRRPAPARSVLI 95  
 QY 267 HDL--GEPIYIDRAVEVREGADKAFITVADEVPLLDGVFHGKYEKKNLINENASA 324  
 DB 96 AKVYBEPVLIAGVDYLOGCAKTDPPYQALVRRDPTKISSVLNKGDFNFTSSLTGLAL 155  
 QY 325 EHGCFDGRMLDRSDVYLIPNTADVSLIYDTGQYRDEVEVFTIDPKTNOITDPPKLP 384  
 DB 156 RRCYFDANMIKSQLGVAQAQHEAFWIDIDPDSGGRYRFGKVI-----QGSQ 201  
 QY 385 VKELLEQLITVMGKRYNIAQAVRAISNDLIATRYFMVNTETIYFPRRQIONDQVFEQ 444  
 DB 202 IRVDYIQLNVLPEFHGEYTSDELAELNRLRLAATNMFNSV--VWSPD-----FOD 248  
 QY 445 SSSRTEPAQVDSLEPIETVELTDLGILMDSPIEFSSNLIDQLNLVAAKARHLYD 504  
 DB 249 AKRSK-----ILPLD----- 258  
 QY 505 MPDRLVLAINHGDVNSILGRISDAVASAARAILPDESENEVIDLERTALANKRPAD 564  
 DB 259 -----AVTPRENTV----- 269  
 QY 565 VYOSKAVPLVYVASDKPRDQGLGWSGTGRLVTKFEHNLINRDGYQAGAEIRLSED 624  
 DB 270 -----ELGGYATDVGPRLTASMKRPMWNSFGHSLTTTALSAP 308  
 QY 625 KKGVL-YATKPLSHPLNDQLRATLGYOEVFGHSTNGFPLSTLHEISRLIQNGW 683  
 DB 309 EQLLDSTYRPLRLNPLEQYIYLIQGFRR-----IDLNDTNSDTTLNVAREWLSGW 362  
 QY 684 NRTYSRLRYLDLKTQAPRETMODLPYDFVNGKPSQEAAL--AGVAHKTVAIDLVPNR 741  
 DB 363 QRAINLRMSLD-----HFTQGRVTDFTMLLYGVSINIRLRQSGANPVM 406  
 QY 742 GYQRIISLEVSSGLVSDANMAIARAGISGVSEFQDNAYGNSNRHQTGCIQAGYIWSDN 801  
 DB 407 GDSQRIASIDVSTTWCSDVDVFGIFQANWMIIRLGEK-----NREVARGNWGIETNN 459  
 QY 802 FNVVPRRLRPFAGGDSIRGYANDSLSPIDKGYLLGGGOVLAVGTAVYENPKMDLRLAV 861  
 DB 460 FDRVPPSLRPFAGGDSIRKYKFRNISPRDSQGLTGASGLATGSLTYQYNFGRMGAV 519  
 QY 862 FGDIGNAYDKGTNDTKIGAGVGVNRASPVGOVAVDVATVCKEKGNIKLHFFIG 916  
 DB 520 FVDSGAVNNFSSDLKTGAGVGVNRASPVGPIKLDIAAI--GDNEHNGVQFYIG 573

RESULT 4

QY 08XELS PRELIMINARY; PRT; 577 AA.  
 ID 08XELS;  
 AC 08XELS;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Putative outer membrane protein (putative exported protein).  
 GN YTFM OR STM4409 OR STY4768.  
 OS Salmonella typhimurium, and  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_taxid=602, 601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-S.typhi; STRAIN-LT2 / SCSC1412 / ATCC 700720;  
 RX MEDLINE-21534948; PubMed-11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 LT2,"  
 RL Nature 413:852-856(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-S.typhi; STRAIN-C718;  
 RX MEDLINE-21534947; PubMed-11677608;  
 RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaita M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gara P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrett B.G.;  
 RA "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi C718,"  
 RL Nature 413:848-852(2001).  
 DR EMBL: AE008906; AAL23329.1; -  
 DR EMBL: AL627283; CAD06889.1; -  
 DR InterPro: IPR000184; Bac\_surfAg\_D15.  
 DR InterPro: IPR000566; Lipooln\_cytfABP.  
 DR Pfam: PF01103; Bac\_surface\_Ag\_1.  
 DR PROSITE: PS00213; LIPOCALIN: UNKNOWN\_1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 577 AA; 64765 MW; FBA4A190173BFL8C CRC64;

Query Match 10.1%; Score 479; DB 16; Length 577;  
 Best Local Similarity 22.7%; Pred. No. 5e-20;  
 Matches 166; Conservative 104; Mismatches 257; Indels 204; Gaps 17;

QY 210 ANTKALEDITQESAMDLNLSIPLRQTALVAAKAVGYD--IDLSTINSGEVDYII 252  
 DB 22 ANVRLKVEGSGELEKVRQOLSTIQSDEVTPDRFRARVDAIREGLAKAGYEPIKE 81  
 QY 253 SIINSGEVDYIIHDL--GEPIYIDRAVEVREGADKAFITVADEVPLLDGVFHG 310  
 DB 82 DLPPKAGKROVLARVTPGQPVLLIGTEVYLIRKGARTDKYLLALTRP-AITVLNQG 140  
 QY 311 KYETKKNLIENASAHEGYFDGRMLDRSDVYLIPNTADVSLIYDTGQYRDEVEVFTID 370  
 DB 141 DYDFKKSILSVSLARKYTFSEFIKSLGIALGHNQAFWIDVDSGEYRGRGPVTF----- 196  
 QY 371 PKTNQLTTDPPKLPVKRELLQLITVMGKRYNIAQAVRAISNDLIATRYFMVNTETIYF 430  
 DB 197 -----EGSQIRVDYIQLNVLPEFHGEYTSDELAELNRLRLAATNMFNSV--VWAP 243  
 QY 431 EREGIONDQVSEFGSSSRTEPAQVDSLEPIETVELTDLGILMDSPIEFSSNLID 490  
 DB 244 E-----FEKSRKTKILP----- 255

QY 491 KLNLYAKARHLYPDMPDDRVLAINHDDGVNRSILGRISDAVSAVARAILPDESENEVIDL 550  
Db 256 -----LKVGSPTTEMTIT--- 268  
QY 551 PERTALANKRTPADYQSKVPLVYFVASDKPRDGOIGLWGSDDTGRLYTKFEHNILNR 610  
Db 269 -----ETGVGYSTVDGPRVKASMKKPMNNS 293  
QY 611 DGYOAGELRLSEDKKGYKL-YATKPLSHPLNDQRLATLGOQEVFGHSTGFDLSTRTL 669  
Db 294 YGHSLLTSTSTISAEQVYDEFTYKMPKLNPLEQYLYLVGGFKRT---DLNDDTDDSTTL 349  
QY 670 EHEISRSITONGWNRRTYSLRYRLDKLTQAPPETWODLPVDFVNGKPSQEQALL--AGVA 727  
Db 350 --AVSRYWDLSSSGQRAINLRWMSFD-----HFGQGNVTNTTMTLFTYECVM 391  
QY 728 VHKTVANLVNPMKGYQORISLEVSGSLVSDANMAIARAGISGVYSGFDNAVGSNRAHQ 787  
Db 392 ISRTSRKGLMPTWGDORSYSDYSNTAMGSDYDFSVLQANWMIRTLDR-----HR 444  
QY 788 MTGGIAGYIWSDNFNHVPRLRPFAGDOSIRGVAHDSISPIDSKGYLTGGVYLVAGTA 847  
Db 445 FVMRANLGMLETODFDKVPPLDRPFAGGDSIRKGYTKTSISPDSDGNLKGASKLATGSL 504  
QY 848 EYNEFMKDLRLAVFGDIGNAYDKGFTNDYTKIGAGVGRVAPGYQVRYD--VATGYKEE 905  
Db 505 EYOYNTYTGKMGAVFVDSGBAVSDIRSRDKTGTVGVGRVAPGYPKLDPFAPVGDKE 564  
QY 906 GNPYKHLHFFIG 916  
Db 565 HG---LQFTYG 572

## RESULT 5

Q9KP31 PRELIMINARY: PRT: 582 AA.  
AC 09KP31: 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)  
DE Hypothetical protein VC2548.  
GN VC2548  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN:EL TOR N16961 / SEROTYPE O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,  
RA McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
RT cholerae.";  
RL Nature 406:477-483(2000).  
DR EMBL: A0004323; AAF95689.1; -  
DR TIGR: VC2548; -  
DR InterPro: IPR000184; Bac\_surfag\_D15.  
DR InterPro: IPR000566; Lipoclin\_cytfABP.  
DR Pfam: PF01103; Bac\_surfactin\_Ag\_1.  
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN\_1.  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 582 AA: 65644 MW: E96946FB85F1C426 CRC64;

Query Match 9.6%; Score 454; DB 16; Length 582;  
Best Local Similarity 22.4%; Pred. No. 1.5e-18;

Matches 164; Conservative 104; Mismatches 237; Indels 228; Gaps 20;  
QY 211 NIKAALEDITQESAMDLNGSI---PLRLQATALVAAVAVY--DIDSLIRNS----- 258

Db 48 NVEAYLSI---AADYSTSLRFQSLERSMTEALNALGYHPSIDYFSEDNQLRAAV 104  
QY 259 -----IGEVNVIITHIDGEPEYIDYRAVEVNGEADKAFFTVADVEPLIGVFHHGK 311  
Db 105 TLGEVTRLSEVDIVL-----RGEAGGRDIFQRLRRSGLRVADAPLHNSL 148  
QY 312 YETKNLLENASAEHGYEDGRWLDRSVDVILPDNTADVSLIYDTGTGYRPFDEVEFTIDP 371  
Db 149 YDNLSKSGIRNALQKGYNGDFQASRLLEVPELNOARVILHFDSGIRLF----- 198  
QY 372 KTNQTLTPDKLPVKRELLEOLLTVNMGAYNLAVALSNDLITATRYFNWVTEIYPE 431  
Db 199 --GATTEGSGIDENRW--SLRPFKQGEPLYVSQGFENQNLSTDMFSSVPE--PD 251  
QY 432 REQIONDVFSFGSSSSSTPEPAQVDESTLEBEVIEVELTDGLMDISPIERSASNLQDK 491  
Db 252 LSQI----- 255  
QY 492 LNLVAAKARHLYPDMPDDRVLAINHDDGVNRSILGRISDAVSAVARAILPDESENEVIDLP 551  
Db 256 -----DEG----- 258  
QY 552 ERTALANKRTPADYQSKVPLVYFVASDKPRDGOIGLWGSDDTGRLYTKFEHNILNRD 611  
Db 259 -----RELPIKVTYLAPOARNQLETGIGYSTVDYGVRSGLKMKKPMWNSQ 301  
QY 612 GYOAGELRLSEDKKGYKLATKPLSHPLNDQRLATLGOQEVFGHSTNGFD--LSTRTLE 670  
Db 302 GHSFDSFSSTISPEQTTTACYKIPLEDALNEYRYIOYGMK-----LDRDRTSLE 352  
QY 671 HEIS--RSITONGWNRRTYSLRYRLDKLTQAPPETWODLPVDFVNGKPSQEQALLGVAV 728  
Db 353 SNLSLEHHQDLDGGMHRTVIRYLLENYR-----QGIQDD-----NSQFLPLPGMTY 398  
QY 729 HKT-VADNLVNPARGYRQRYSLLEVSGSLVSDANMAIARAGISGVYSGFDNAVGSNRAHQ 787  
Db 399 TRTRTSNSGILTWGDKQTITTEYGDPALSTETRYRLQTGSSWMTYRAN-----HR 451  
QY 788 MTGGIAGYIWSDNFNHVPRLRPFAGDOSIRGVAHDSISPIDSKGYLTGGVYLVAGTA 847  
Db 452 ALVRVYDGANLVDEFQDLSPLSLRFAGGNLNLGKGYKTSIPDASGALTGAAYIATSSI 511  
QY 848 EYNEFMKDLRLAVFGDIGNAYDKGFTNDT---KIGAGVGRVAPGYQVRYVAGV-K 903  
Db 512 EYOYRLTGMWMAAMFMDVGDAF-----NNPEWKKGVGTGIRNISVGPRLDFAPWGLDA 566  
QY 904 EGNPKHLHFFIG 916  
Db 567 APDEKIKHLFTIG 579

## RESULT 6

Q9EY32 PRELIMINARY: PRT: 593 AA.  
AC 09EY32: 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
DE Putative outer membrane protein.  
OS Xanthomonas oryzae pv. oryzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
OX Xanthomonas.  
OX NCBI\_TaxID=64187;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Goel A.K., Rajagopal L., Nagesh N., Sonti R.V.;  
RT "Xanthomonadin biosynthetic gene cluster from Xanthomonas oryzae pv.  
RT oryzae.";  
RL Submitted (MUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY010120; AAC38834.1; -  
SQ SEQUENCE 593 AA: 66684 MW: 0FF8A8C135C4F91A CRC64;



Db 473 LRWHEGLGASRLIRG-----EAGTTWASDLVAMPSPSLREFAGVNSIRGVAFREVG 525  
QY 829 PISDK--GYLTGGVLAAGTAETAEYNEFM-KDLRLAVFGDIGNAYDKGTNPTKIGAGVY 885  
Db 526 PRYAKPDAFALGAKHVFSAAGAEFEYYKGGPFGCAVFDSDSAPNR--YPMHHTGVGIGL 583  
QY 886 RMASPVGVQVRDVATGKVEEGNPILKHEFFIG 916  
Db 584 KYRSPGVGVVDIARGLNSPDSKIOLYIDIG 614

RESULT 8  
QBRHT4 PRELIMINARY: PRT: 678 AA.  
ID QBRHT4  
AC QBRHT4  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Outer membrane protein.  
GN FN1911.  
OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteria; Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25586;  
RX MEDLINE=21886394; PubMed=11889109;  
RA Kapatala V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,  
RA Larssen N., Kyrpides N., Malunas T., Pusch G., Haselkorn R.,  
RA Fongsteijn M., Kyrpides N., Overbeek R.;  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
nucleatum strain ATCC 25586.";  
RL J. Bacteriol. 184:2005-2018(2002).  
DR EMBL: AEO10493; AAL94010.1;  
KM Complete proteome.  
SQ SEQUENCE 678 AA: 76229 MW: FB3579AFD639A74B CRC64;

Query Match 5.6%; Score 262.5; DB 16; Length 678;  
Best Local Similarity 22.6%; Pred. No. 4e-07;  
Matches 162; Conservative 111; Mismatches 265; Indels 179; Gaps 36;

QY 311 KYETKKNLE-NASAEHGFDGRMLDRSDVILPDNTADVSLYDTGTQYRFD---EYV 365  
Db 33 KESTEALLADFNALKETGYFE-----DVILQ-----PVSVDGVRIVDVKEKNV 78  
QY 366 FTIIDPK---TNOLTTDPKLPY-----KRELLLEOLLIVNM--GEAYNLQAVRAL 410  
Db 79 VDLKRGVAINTLRDSDSYLSSVKTGMKRVTTSELDITOLKACGYFSRSVEDA 138  
QY 411 SNDLIATRYFNMYNTEIVFEREQIOND--VSFEQSSSRTEPAQVDESLTSLVETVE 468  
Db 139 QRRLLATGKFESEVRPD-----AQVANGKMAISFE-----VVEN--PIKSVI 178  
QY 469 LVDGIIMDISPI--EBSA-----SNLIQDK--LNLVAKARHLQYMPD-----DRV 510  
Db 179 ITCNNNTIPTSTIMSELTTKPGSVQYNNLREDRDKILGLYQAGYTLVNTIDKSTDENGT 238  
QY 511 LAIINHDDGVNRSITGR-----ISDAVASAARAILPDESE---NEVIDLPERTA 555  
Db 239 LHIISVEGIVRRLEVKKMTVKOKGNRRTPNDVLTAKTDVYIDREIIOPKIFNVKEYDA 298  
QY 556 LANRRTPADV-----YQSKVP-----LYFVASDKPRDQIGLGMGSDGTSLVTKF 603  
Db 299 TVDNLMLRGLIFKNVKEYEASIFQDPEGIDILILIDEDRAELQAGVAVSSETE-----F 352  
QY 604 EHNILNRDGYQAGAE---LRLSEDKRGVLYATKPLSHPLDQLERATLGYOQEVFGHST 659  
Db 353 LCTLSLKDSNMGRKMQQGFTEFKSNKNTYGFALDGYDPWIMKDTORVSMGW-----GAGR 407  
QY 660 NCFDLSTRLEHEIS---RSIIQNG-GNNRRTYSLRYRLDKLTKTO----- 699

Db 408 TSYGDESDILFEHIDTIGFRTNIGKLGKNFTLSLGTWEYIEKEHDEGKLRQANNGKWY 467  
QY 700 -APPTNODLPRDVPNGKPSQEALLAGVAVKTVADNLVNPGRKGRORXSLEFVSSGLVS 758  
Db 468 YKEKNKRE--IEGVDDKYLWMSITPYISY--DTNNNNLNPSTSGFYGFGVEAGHAGGK 523  
QY 759 DANMAIARAGISGVYS--FGDNAYGNSRAHOWMTGIGQAGYIMSDNFNVPRLRFAGAD 816  
Db 524 SGNENATLLELRTYKIGLFGKNFIF-----AKYVGVAI-----NNKESQKFWGG 571  
QY 817 QSRIGYAHDSLSPISSDKGYLTGGVLAAGTAETAEYNEFMKDLRLAVFGDIGNAY-----DK 871  
Db 572 NSLRGY-----DGGFFKGSQKL-VATIENTRQLNDIIGLVFADAGRAMKQNGRDP 621  
QY 872 GFTNDT-----KIG--AGVGYRAMSPVQVQVRDVA--TGKVEEGNPILKHEFFIGTPF 919  
Db 622 SYTRDSNRFGHNIGTTAGVIGIRLNPILPGLRFDGWPVGNKMDDGKMFYFMGGSF 678

RESULT 9  
P95359 PRELIMINARY: PRT: 792 AA.  
ID P95359  
AC P95359;  
DT 01-MAY-1997 (TREMBlrel. 03, Created)  
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Outer membrane protein.  
GN OMP85.  
OS Neisseria gonorrhoeae.  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=485;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FA19;  
RA Manning D.S., Reschke D.K., Judd R.C.;  
RT "Omp85 proteins of Neisseria gonorrhoeae and Neisseria meningitidis  
are similar to Haemophilus influenzae D-15-Ag and Pasteurella  
multocida Oma87.";  
RL Microb. Pathog. 23:0-0(1998).  
DR EMBL: U81959; AAC17600.1;  
DR InterPro: IPR000184; Bac\_surfAg\_D15.  
DR Pfam: PF01103; Bac\_surface\_Ag\_1.  
SQ SEQUENCE 792 AA: 87868 MW: 90E32D24AA0513D8 CRC64;

Query Match 5.8%; Score 259.5; DB 2; Length 792;  
Best Local Similarity 20.6%; Pred. No. 7.9e-07;  
Matches 189; Conservative 119; Mismatches 355; Indels 255; Gaps 37;

QY 116 TPLSLELFAQESTEMGINDND-----YIPEYQGEQPNSEVVPPLFEPEKQLIKRLY 169  
Db 16 SPLAFADFTIQDIRVEGLQRTPESTVFNYLPVKVGDYTN-----DTHGSAIIRSLY 66  
QY 170 ARLENDGV-----NKPRLKAKFYQSSQ--SGETSAGISSHOKTEPY 209  
Db 67 ATGFEFDVREYETADGILLTLVIVCPTIGSLNTIGAKMLQMDALKKNLESGLAQSOYFNQ 126  
QY 210 ANIKALEDTQSSADNLSIRPKOTFALYAARANGYIDISITNSIGEVDTIIDL 269  
Db 127 ATLNOQAVAGKEEY---LGRKLNIOITTPKVTKLARNRVDIDITIDEGKSAKITDIEFE- 182  
QY 270 GEPVYIDYRAVEYRGAGADDAKFTVADEVPLLIGVFNHGKYTEKKNLENSAHEGYF 329  
Db 183 GNGVYSDRKIMR-----QMSLTBEGIGMTWLTNRSDRFQKFAQDMEKVTDYQNNNGYF 235  
QY 330 DGRWLDRSVDVILPDNTADVSLIYDTGTQYRFDDEVVFTIDPKTNOLTTDDPKLPVKREL 389  
Db 236 DFRILDTDIQTNEDKTRQRTIKITVHEGGRPMGKV---SIEGDTNE-----YPKAE 283  
QY 390 LRGQLLVNMGGEAYNLQAVALSNDLATRYFNMYNTEIVFEREQIONDQVSFEQSSSR 449  
Db 284 LEKLLTKKPKGYERQOMTAVILG-----ETONRGSAGVAYS-- 320  
QY 450 TEPQVDESLTLEPVIETVELTGGIIMDISP-----IEFSASNLIQKLNLYAKAKAH 501

Db 321 -----EISVQPLPNAAGTVDVFLHIEFGKRIYVNEHIITGNKTRDE--VVRRLRQ 371  
QY 502 LYDMPD-----RVLAINHDDGVNRSILGRISDAVAVARAILPDESENEVIDLPE 552  
Db 372 MESAPYDTSKLQSKERVELLGFEDNVQ-----FPAVPLAGIPDK-----VDL-- 414  
QY 553 RTALARKRPADVYQSKKVPYVVASDKPRDQIGL--GMSDPTGRVLTKEEHLINR 610  
Db 415 NMSLTERST-----GSLDLSAGVQDTP--LVMSAGVSQDNL 449  
QY 611 DGVOAGAEILRSEDKKGKVLATKPLSHPLNDLRATLGVOQVEFGHSTNGFDLSTTLE 670  
Db 450 FGTGKSAALRASRK--TTLNGSLSTDPYFTADGVSLG--DLYGAEPKRAKASTSVKQ 505  
QY 671 HEISRSIIIONGWNRTYSLRRLDKLTQAPPE-----TWODLP--VDFVN--GK-PSQ 719  
Db 506 YKTTTA-----GGGVRRGIPTEYDRVNFGLAHLVNTYNKAPKRAADIKQYKTDGA 561  
QY 720 EALLAGVAAHKTY-----ADNLVNPGRGYRQRYSLVEGSSGLVSDANMAIARAGISGVY 773  
Db 562 DGSFKGLYKGYWGMRNKTDLSALMPTRGY-----LTGVNAEIALPGSKLQY 608  
QY 774 SFGDNAYGSRBAHQMG-----GIOAGYIWSDNFNHVPYRLKFFPAGDQ 817  
Db 609 -----YSATHNQTFWEPPLSKFTTLMLGEGVLAGGY--GRTKEIPFENFEGGGLG 657  
QY 818 SIRGVANDSLSPISDKGYLTGGQVLAVG--TAEVNFEF-----KDLRLAVGDI 865  
Db 658 SVRGYSSTGLCP--KVYDEYGEKISYGKKNKANVSALLFPMPGAKADARTVRLSLFADA 714  
QY 866 G-----NAVDKGFTNDTKIGAGVYKMASPVGQVRVDVATGVK 903  
Db 715 GSWMDRTYTAENGNNKSVYSENAKSTFTNELRYASAGAVTWLSPLGMPKFTYAVPLK 774  
QY 904 E--EGNPRIKLFPIGTFP 919  
Db 775 KKEPEIDQRFQGLGTFP 792

RESULT 10  
O9K1H0 PRELIMINARY: PRT: 797 AA.  
AC O9K1H0: 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE Outer membrane protein Omp85.  
GN NMB0182.  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID:491;  
RN [1]  
RP SEQUENCE FROM N. A.  
RC STRAIN-MC58 / SEROGRUP B;  
RX MEDLINE-2017575; PubMed-107110307;  
RA Tetteil H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
RA Eison J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,  
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,  
RA Cotton M.D., Ulteback T.R., Khouri H., Qin H., Vamathevan J.,  
RA Gill J., Scarlato V., Masiugnani V., Pizsa M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,  
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
RT MCS8".  
RL Science 287:1809-1815(2000).  
DR EMBL: AE002375; AAF40639.1; .  
DR TIGR: NMB0182; .  
DR InterPro: IPR000184; Bac\_surfAg\_D15.  
DR Pfam: PF01103; Bac\_surface\_Ag\_1.  
KW Complete proteome.  
SO SEQUENCE 797 AA; 88436 MW; B68BFC5A3D22E8 CRC64;

Query Match 5.5%; Score 259; Db 16; Length 797;  
Best Local Similarity 20.5%; Pred. No. 8.6e-07;  
Matches 191; Conservative 121; Mismatches 342; Indels 278; Gaps 40;  
QY 116 TPVLSLELFAQESTEMKINRD-----YIPYOGEDPNSEVVVPPVLEPEKELIRLY 169  
Db 16 SPLADFTIODIRVELQRTPESTVENYLPVKGDHYN-----DTHGSAIIXSLY 66  
QY 170 ARLEFNDGVNVP--RLKAFYQSSQSETSAGSSHQKTEPYANIKALED--ITQESA 224  
Db 67 ATGFEDVRYETADGQLLTVIERPTIGSLNITGAKMLQDA--IKKNLESFGLAQSY 123  
QY 225 MD--LNGSIPRLQ-----TALVAARAVYIDLSIIRNSIGEVNIIHDL 269  
Db 124 ENQATLNQAAVAGLKEEVLGKGNLIQITPVYTKAARRVVIDITIDGSKAKITIDIEF- 182  
QY 270 GEPVYIDYRAVEYRGEGADKAFITVADEVPLIGDV-----FHNGYETKRLIE 320  
Db 183 GNOVYSDRKLMR-----QMSLTEGGIWTWLTNSNQFNEQKFAQDMKVT 226  
QY 321 NASAEHGFEDGRMLDRSDVILPDNTADVSLIYDTGYRFEDEVFTTIDPKTNQLTDP 380  
Db 227 DFTQNNGFEDFRILDDIQTNEDKTKOTIKTYHEGGFRMGKY--SIGDTNE----- 278  
QY 381 DKLPVKRELLEQLITVNMGEAYNLQAAVRLSNDLIATRYFNMVTEIYPEREQIQDQV 440  
Db 279 --VPRAELEKLLTMKRGKMYERQOMTAVLG-----EIQRMG 313  
QY 441 SFEQSSSRTPEAQVDESLTEPIYETVELTDLMDISP-----IEFSASNLIDKLT 492  
Db 314 SAGYAVS-----EISVQPLPNAETKTVDVFLHIEFGKRIYVNEHIITGNKTRDE- 363  
QY 493 NLVAAKARHLYDMPD-----RVLAINHDDGVNRSILGRISDAVAVARAILPDES 543  
Db 364 -VVRRLRQESAPYDTSKLQSKERVELLGFEDNVQ-----FPAVPLAGIPDK- 411  
QY 544 ENEVIDLPERTALANRTPADVYQSKKVPYVVASDKPRDQIGL--GMSDPTGRVLT 601  
Db 412 ----VDL--NMSLTERST-----GSLDLSAGVQDTP--LVN 440  
QY 602 KFEHNLINRDGYOAGAEILRSEDKKGKVLATKPLSHPLNDLRATLGVOQVEFGHSTNG 661  
Db 441 SAGVSODNLEFGTGSALRASRSK--TTLNGSLSTDPYFTADGVSLG--DLYGAKFDP 496  
QY 662 FDLSTRTLEHRSIIIONGNR-----YSLRLDKLTQAPPTWODLP--VDFV 713  
Db 497 RKASTSIKQYKTTTA--GAGIRMSVPYTEDRVNFGVLAHLVNTYNKAPKRAHYADFT 552  
QY 714 N--GK-PSQALLAGVAAHKTY-----ADNLVNPGRGYRQRYSLVEGSSGLVSDANMAI 764  
Db 553 KKYGTGIDGDSFGKGMLYKCTVGMGRNKTDLSALMPTRGY-----LTGVNAET 599  
QY 765 ARAGISGVYSGDNAYGSRBAHQMG-----GIOAGYIWSDNFNHVPYR 808  
Db 600 ALPGSKLQY-----YSATHNQTFWEPPLSKFTTLMLGEGVLAGGY--GRTKEIPF 648  
QY 809 LRFAGGDSIRYANDSLSPISDKGYLTGGQVLAVG--TAEVNFEF-----KD 866  
Db 649 ENFYGGGLSVRGESGTLGP--KVYDEYGEKISYGKKNKANVSALLFPMPGAKADART 705  
QY 857 LRLAVDIGNMAD-----KFTNDTKIGAGVYKMASPVGQVRVDVATGVKMS 889  
Db 706 VRLSLFADGGSVDGKTYTDNSSATGKRVQNTYGCNTHKSTFTNELRYSAGAVTWLS 765  
QY 890 PVGQVRVDVATGVKE--EGNPRIKLFPIGTFP 919  
Db 766 PLGPMKFSYAYPLKKKPEDEIQRFQGLGTFP 797

RESULT 11  
O9UXJ1 PRELIMINARY: PRT: 797 AA.

AC 09JXJ1;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE Outer membrane protein OMP85.  
 OS OMP85 OR NMA0085.  
 GN Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_Taxid=56599;  
 RN 11;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2491 / SEROGROUP A / SEROTYPE 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Jagers K.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S.,  
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrall B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis 2491."  
 RL Nature 404:502-506(2000).  
 DR EMBL: AL162752; CAB83401.1;  
 DR InterPro: IPR000184; Bac\_surfAg\_D15.  
 DR Pfam: PF01103; Bac\_surface\_Ag; 1.  
 DR Complete proteome.  
 SQ SEQUENCE 797 AA; 88404 MW; 65DE47E0C9E1D1F CRC64;

Query Match 5.5%; Score 259; DB 16; Length 797;  
 Best Local Similarity 20.5%; Pred. No. 8.6e-07;

Matches 191; Conservative 121; Mismatches 342; Indels 278; Gaps 40;

QY 116 TPLSELEFAESTENGINPND-----YIPEVGEOPNEVVVPTLEPEKGLIKRLY 169  
 DB 16 SPLALADFLIDIRVGLQRTERTFPNLPVAVGDTYN-----DHGSAILIKSLY 66  
 QY 170 ARLFNDGVKVP---RLAKFYOSSGSETSAIGSSHOKTEPYANKALAE--ITQESA 224  
 DB 67 AGCFEDVAVETADGQLLTIVIERPTIGSLNITGAKMLQND---IKNLESGLAQSOY 123  
 QY 225 MD---LNGSIPRLRO-----TALVAVARAVGYDIDISTIRNSIGEVYIHD 269  
 DB 124 FMOATLNOAVAGLKEEYLCGRKLNIOITPKYKLARNRVIDITIDEKSAKITDIEFE- 182  
 QY 270 GEPYIDYRAVEVRGEGADKAFITVADEVPLIGDY-----FHGKYEKKNL 320  
 DB 183 GNGVYSDRKLAM-----QMSLTGEGIMTWTLSNOFNOKFAQDMKVT 226  
 QY 321 NASAHEGTFDGMRLDRSDVILPDTADVSLIYDTGYRFEDEVVFTIDPTKNTLTD 380  
 DB 227 DFYQNGGYFDFRILDTDIQTNEDKTKOTIKITVHGGFRMKV---STEGDTNE----- 278  
 QY 361 DKLPRKRELLBOLLVNNGEANLQAVRALSNDLIATRTFNVMNTEIYFPEREQIONDV 440  
 DB 279 ----VPKAELEKLLTMKKGKWERQOMTAVLC-----EIONRMG 313  
 QY 441 SFEQSSSRTEPAQVDESTELEPIETVELTDLGILMDISP-----TEFSASNLIOK 492  
 DB 314 SAGAVYS-----ELSVQPLPAETKTYDVFVLAIEGRKIYNEIHTGNNTKTRDE- 363  
 QY 493 NLVAAKARHLTYDMPD-----RVLAINHDDGVNRSIIGRISDAVAVARALLPES 543  
 DB 364 -VVRRELLQMESAPDYTSKLQRSKERVELLDYFDNVQ-----PDVAPLAGTTPK- 411  
 QY 544 ENEVIDLPERTALANRKTTPADYOSKVPYLVFAASDKPRDQIGL--GWSGDTGRLVT 601  
 DB 412 ----VDL--NNSLTERST-----GSLDLSAGWQDTG--LVM 440  
 QY 602 KFEHNLINRDYQAGAEIILSEDKKGVLYATKPLSHPLNQLATLGQOQDEYGHSHNG 661  
 DB 441 SAGVSDNLPOTGSAALRASRSK--TTLNGSLSTFDYPTADGVSGLY--DYVGKADP 496  
 QY 662 FDLSTRLEHEIRSIIONGMNR-----TYSRLRYLDDKLTQAAPETWQDLP--VDVF 713

DB 497 KKAISTIKOYTTTA---GAGIRMSVPVTEYDRNFEGLVAHEHLTVNKNRAPHYADFI 552  
 QY 714 N-GK-PSOALLAGVAVHKTV-----ADNLVMPMRGYRORYSLEVGSSGLVSDANAI 764  
 DB 553 KKYKRTGCTGDSFGKGLWKYGVGGRNKTDSALMPTKRY-----LGVNAEI 599  
 QY 765 ARAISGVYSGDNAVGSNRAHOMTG-----GIQAGYIWSDNFNHVPYR 808  
 DB 600 ALPGSKLOY-----YSATHNQTFEPLSKTFTLMLGVEVGIAGY--GRTKEIPF 648  
 QY 809 LRFAGDGSIRGVAHNDSLSPISDKGLTGOVLAVG---TAEINYEPM-----KD 856  
 DB 649 ENFYGGGLGYSRGYESGTLG---KYVDEYGEKISYGNKKANYSABELPMPGAKDART 705  
 QY 857 LRLAVFGDIGNAYD-----KGFNDTKIAGGVYRMAS 889  
 DB 706 VRLSLFADAGSYMDGKRYTDNSSSATGGRONITGAGNTHKSTFTNELRYSAGGAVTWLS 765  
 QY 890 PVGOVRVAVATGYKE--EGNPIKLHFFIGTPE 919  
 DB 766 PLGPMKFSYAVPLKKRPEDEIQRFOGLGTF 797

## RESULT 12

08VQ24 PRELIMINARY; PRT; 798 AA.

AC 08VQ24;  
 DT 01-MAR-2002 (TReMBLrel. 20, Created)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Omp85.  
 GN Omp85.  
 OS Bartonella henselae (Rochalimaea henselae).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Bartonellaceae; Bartonella.  
 OX NCBI\_Taxid=38323;  
 RN 11;  
 RP SEQUENCE FROM N.A.  
 RA Zimmermann R., Augustin K., Schaal K., Sander A.;  
 RT "Cloning, nucleotide sequencing, and expression of a hemin-binding  
 protein of Bartonella henselae".  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF461795; AAL65374.1;  
 DR InterPro: IPR000184; Bac\_surfAg\_D15.  
 DR Pfam: PF01103; Bac\_surface\_Ag; 1.  
 SQ SEQUENCE 798 AA; 88985 MW; 96A3E9C9066B7084 CRC64;

Query Match 5.4%; Score 254; DB 2; Length 798;  
 Best Local Similarity 21.9%; Pred. No. 1.7e-06;

Matches 203; Conservative 136; Mismatches 352; Indels 236; Gaps 52;

QY 96 SRIGQSPPLGDMSVIETPTPL-----SLEELFAQESTEMGINPNDYIPEYGEOPNSEV 151  
 DB 5 SRFLLAASVLYGMKVIAPTTAFVSIAMVEEYQAVNSIEVHKNFV--SAGVYVNDNTDI 63  
 QY 152 VVPPTLEP-EKPLIKRLYA--RLFND-GVNKY-PRU-----KAKFYOSSGSETS 197  
 DB 64 KYGKSFSSGDVDVAVKRLFALGLFDYKINGDLYLVKYEYVNVNOLYFGNKSLLDP 123  
 QY 198 AIGS--SHOKTEPYANITAALEDITQESAMDINGSIPRLROTALVAVARAVGYDILSII 255  
 DB 124 DLKRFISLKPNEPENSAR-----LSADVNIARE---AYKTGVNDIAVITQ 166  
 QY 256 RNSIGSEVDV-IIHDLGEVYIDYRAVEVRGEGA--DQKAFITYADEVP-----LLIGVF 307  
 DB 167 TINLCKGVNVVENVNVEBGRRTKISNITFKGNHAGSSRLRDYISTKPSGILSLLRGDY 226  
 QY 308 HHGYETKKNLEENASAEHGYFDGRWMLDRSDVILPDTADVSLIYDTGYRFEDEVV 367  
 DB 227 SEERLADEALRPFYRYRGYADRFVYSSKAVFDEARAAVEIDVLDGAVKTKIGDV--- 283  
 QY 368 TIDKTKNOLTTDPPKLPYKRELLBOLLTVNNGEAY--NLQAVRALSNDLIATRTFNVMN 424

```

Db 284 -----QISDIDGIDV---QSMKGILKTRSGDYYSAENIEQSVAILNNKAADSGYAFAK 334
QY 425 TELVFE-REQIONDQVS-----FEQSSSRTEPAOV--DESTLEPV-ETVELDGLIMD 476
Db 335 VE---PRGNNDLANHTISILYNEQGPRAVQREIENGNEKTRDYIRREIDLEGAYN 391
QY 477 ISPIE-----FSASNLIOQKLNLAAKARHLVDPDDRVLAINHDDGVRSLG 525
Db 392 QTLVQRAKRRLRESLGFPAVN-----ISWPT-----DQPDQILLVVD-----VLEAPTG 436
QY 526 RIDDASAVARALLPDESENEVIDLPERTALANKRTPADVYQSKKVPVYFVASDKPRDG 585
Db 437 DLFSFGYTTGTSPGVSLF--VSVTERN-LGGRGQ-----YV----- 471
QY 586 QIQLGMSDGTGRLVTFEHNILNRDGYQAGAELRLED-----KKGVKLYATKP----- 635
Db 472 RLGLGAGQEKSR-----ANNLSFVDPYFLGRLSAGVDLFRSTYRADKAYDVQRQGSGL 525
QY 636 -LSHPLNDOLRATLGY--OOEV-FGHSTNGFPLSTRT--LEHEISRSII--QNGW 683
Db 526 RFAYPINDOLSANLAYSYIQEYDFGK--YDLSKETDIRELYGKYSAGALVQAQHSFW 582
QY 684 NR---TYSLRFR-LDKIKTQAPPETWODLPDYVYNGKPSQEALL--AGVA-VHKTVADEL 736
Db 583 KRSSISYGLYNTIDMKN--PHDGMVVRVLOEYAGLGNKAKFLKTKGKAMMYKTLSL-- 638
QY 737 VNMGRGYRQRYSLVSGSSGLVSDANNAIARAGISGYVSPDNMAYGSRARAHOMTGIOAGY 796
Db 639 -----RMDLVGLSLFSG-----GGY 652
QY 797 IMSDNFNHVPYRLRFAGGQDQIRGYAHDSLSP--ISDKG--YLTGGOVLAVGTAEYNYE 852
Db 653 IHEIGREGVRIEFMFKINTDM-IRGFYNGICGRQVSNKEVYFLGTTTNNMTAEVQFP 711
QY 853 FM-----KDLRLAVFGDI-----GNAYDKGTTNDTKI-----GAGGVNRMASPVGYV 894
Db 712 IPVPEGLCFRGLFADVAATLYGSNNKPVQGETPYTNIKSAMRSSAGVSLMDSPFGPL 771
QY 895 RVDVATGV-KEEGNPI-KLHFFIGTPE 919
Db 772 RFDYAMPRIKKQEGDRLQQLNFGISTKF 798

RESULT 13
C030912
ID 030912 PRELIMINARY: PRT: 797 AA.
AC 030912:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Outer membrane protein Omp85.
GN OMP85.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HH:
RT "Omp85 proteins of Neisseria gonorrhoeae and Neisseria meningitidis
RT are similar to Haemophilus influenzae D-15-Ag and Pasteurella
RT multocida Oms87."
RL Microb. Pathog. 23:0-0(1998).
DR EMBL: AF021245: AAC17599.1:
DR InterPro: IPR000184: Bac_surf9-D15.
DR Pfam: PF01103: Bac_surf9-Ag; 1.
SQ SEQUENCE 797 AA: 88539 MW: CP911BSF70B999CF CRC64;

Query Match 5.4%; Score 253; DB 2: Length 797;
Best Local Similarity 20.4%; Pred. No. 1.9e-06;
Matches 190; Conservative 121; Mismatches 343; Indels 278; Gaps 40;

```

```

QY 116 TPISLEELFAOESTEMGINPND-----YIPEYQGEQPNSEVVPPTLEPEKPLIKRLY 169
Db 16 SPLAFAPFTIQDINVEGLQRTPESTVENYILPVKAGDTYN-----DTGSAIITSLY 66
QY 170 ARLEFNDGVNKP--RLKANFYQSSQSGEYSAIGSSHQKTEPYANIKAALED--ITQESA 224
Db 67 ATGFDDVRVETADGQLLLTYIERPTIGSLNITGAKMLQND--IKKNLESFGLAQSY 123
QY 225 MD--LNGSTPLRQ-----TALVAARAVGYIDLSITINRSIGEVDYIITHDL 269
Db 124 FNAQTLLQAAVAGLKEEYLGKGLNIQITPPKVTLLARRRVDDITIDGSKSAKITDIEFE- 182
QY 270 GEPYIDYRAVEVGEQADDKAFTTVADEVPLIGDY-----PHHGKYEKKMLIE 320
Db 183 GNOVYSRKLNR-----QNSLTGEGIWTMLTNSNOTEOKFAODMEKVT 226
QY 321 NASEHGYFDGRWIDRSVDVILPDNTADVSLIYDTGYRDEVFYFTIDPKTNQLTTDP 380
Db 227 DFYQNNGYFDFRILDTDIQNEDEKTKOTIKITYHEGGRFRMGKV--SIEGDYNE----- 278
QY 381 DKLPRKELLEQLTYVMGEAYNLQAVRALSNDLIATRYENMNTETVFPREQIONDQV 440
Db 279 ---VPRAELEKLLTMKPGMYERQOMTAVLG-----EIQNRMG 313
QY 441 SFEQSSSRTEPAQVDESTLEPIETVELDGLIMDISP-----IEFSASNLIOQKL 492
Db 314 SAGYAYS-----EISVQPLPNAETKTVDVFLHIEPRKLYVNEIHITGNKTKDE- 363
QY 493 NLVAAKARHLVDPDD-----RYLAINHDDGVNRSLGRISDAVASARAILPDES 543
Db 364 -VYRRELQKESAPYDTSKQQRKEREVELLGYFDNQ-----FDVAVPLAGTDDK- 411
QY 544 ENEVIDLPETALANKRTPADVYQSKKVPVLYFVASDKPRDQIGL--GNQSDGTRELVT 601
Db 412 ---VDL--NMSLTERST-----GSLDLSAGWVODTG--LVW 440
QY 602 KFEHNILNRDGYQAGAELRLEDKKGKLYATKPLSHPLNDOLRATLGYQOEVFGHSTNG 661
Db 441 SAGVSODNLFCTGKSALRASRSK--TTLNGSLSTFDYPTADGVSLGY--DVGKRAFDP 496
QY 662 FDLSTRTLEHEISRSIIONGWNR-----TYSLRRLDKLKTQAPPETWODLP--VDFV 713
Db 497 RKASTSIKQYKTTTA---GAGIRMSVPVTEYRVNGLVAEHLVYTYKAKAKRHADFI 552
QY 714 N--GK--PQOEALLAGVAVHKTV-----ADNLVPMKGYRORYSLEVSGSSGLVSDANMAI 764
Db 553 KKYGKTDTGDTGSFEGMLYKGTGWCGRNKTDALMPTRGY-----LTGVNAEI 599
QY 765 ARAGISGYVSFGDNMAYGSRARAHOMTG-----GIQAGYIWSDNFNHVPYR 808
Db 600 ALPFSKLQY-----YSATHTQWTFEPLSKTFTLMLGGEVGIAGGY--GRTKEIPIFP 648
QY 809 LRFAGQDSIRGYAHSDSLPSIDKGYLFGGOVLAVG---TAEYNYEFM-----KD 856
Db 649 ENFYGGGLGSVRYEESGTIGP---KYDYEGEKISYGKKNKANVASLELLPMPGADART 705
QY 857 LRLAVFGDIGNAV-----KGFNDTKIGAGVGRMAS 889
Db 706 VRLSLFADAGSVWDCKTYTDNSSSATGGRVQNTYAGAGNTHKSTFTELRLRSAGAGATWLS 765
QY 890 PYGOVAVDVATGVKE--EGNPIKLHFFIGTPE 919
Db 766 PLGPMKFRYAYPLKKRPEDEIQRFQIGTTF 797

RESULT 14
C08XZ13
ID 08XZ13 PRELIMINARY: PRT: 765 AA.
AC 08XZ13:
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Putative outer membrane signal peptide protein.

```





QY 699 QAPPETWQDLPVDFVNGKPSOEALLAGAVAHKIVADNLVNMGRYKORYSLEVGSSGLVS 758  
DB 391 EGA-----ARELSKALVPGFAWTRRDVDDLPFRGRNITLTGAAVAKGLLT 437  
QY 759 DAMMATAIRAGISGVYSGDNAYGNSRAH-----QMTGGIOAGIYMSDNHNHVPRLRFAGG 815  
DB 438 DQSFVNSYSYIRQYVPVQGDIFVARAEALGAVLTGAADG-----VPATLRFRFG 488  
QY 816 DOSIRGYAHDLSL-----SPISDKGYLTGGQVLAFTAENYEFPMKDLRLAVFGDIGNAY 869  
DB 489 TQSIIRGYDFQSIGNDVAGSLTPKFLVLTGG-----VEYQRMFLPMQGAALVFDGTAT 541  
QY 870 DKQFTNDTKIGAGVGVNRASPVGQVRVAVATGYKE 904  
DB 542 DNMWTRRMFNGVGVYMRKSPVPIQLDLAYGIIQ 576  
RESULT 16  
08UC83  
ID 08UC83 PRELIMINARY; PRT; 641 AA.  
AC 08UC83;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DN Hypothetical protein Atu2615.  
GN Atu2615 OR AGR\_C.4742.  
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=176299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21608550; PubMed-11743193;  
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. Sr.,  
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,  
RA Raymond C., Rouse G., Saenphlammachak C., Wu Z., Romero P., Gordon D.,  
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
RA Nester E.W.;  
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
RT C58.";  
RL Science 294:2317-2323(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21608551; PubMed-11743194;  
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
RA Qureshi B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
RA Hounmel K., Gordon M., Vaudin M., Iarchouk O., Epp A., Liu F.,  
RA Mollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,  
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,  
RA Cleo C., Slater S.;  
RT "Genome sequence of the plant pathogen and biotechnology agent  
RT Agrobacterium tumefaciens C58.";  
RL Science 294:2323-2328(2001).  
DR EMBL: AE009208; AAL43596.1; ALT\_INT.  
DR EMBL: AE008175; AAK8337.1; -  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 641 AA; 68281 MW; ECOA1IDA3BDF61E CRC64;  
Query Match 5.0%; Score 236.5; DB 16; Length 641;  
Best Local Similarity 20.3%; Pred. No. 1.3e-05;  
Matches 162; Conservative 107; Mismatches 246; Indels 283; Gaps 31;  
QY 200 GSSHOKTEPYAIKALEIDTQESANDLNGSI-PLRQTAALVAVARAGYVDLSTIRNS 258  
DB 49 GSEPEVEYINVKYA---VTIDAA-DADKSLKSSLENSLLIADKDRASGLGLIIRA 104  
QY 259 ICEVDYIIHDLCEPV-YIDYRAVEVNGEGADDAKFTVADE---VPLLI---GDVFHHG 310

DB 105 RDDDRLLIAALYENARAGGIYNAVAGKNVDLPNPFVDFHSTVPVMTVTPEPKFTIG 164  
QY 311 KYE-----TKKNLNNMSAEHGYFDGRDLDRSVYIIPD----- 344  
DB 165 NVRLLEGDYTGRLNDEYGLIAGDAGSLAIRAGKLLIDDLKAEGRPLAKLTKREAVANHA 224  
QY 345 -NTADVSLYDPTGYQYFEDEVFEFTIDPKTNQTLTTPDPKLPVKREILEQLJTVNMGAYN 403  
DB 225 TNYVDITMAEGGVAPLGNATV-----TGEKIVDGD-----FIRYRSLNGGEYS 271  
QY 404 LQAVRALSNDLIATRYNNMVTIYVPEREQIQNDQVSFEQSSSRPEPAQVDESTLEPV 463  
DB 272 PEKLRKAADRLRLQGVF-----SSLFTRKAGFLARDGTIPL 307  
QY 464 IETVELDGIIMDSPIEFSSANLQDKNLVAAKARILYMPDRVLAIHHDGV----- 519  
DB 308 -TIEVSEG-----KHRYFGVGAQY-----STTEGIGLOG 335  
QY 520 ---NRSIIG-----RISDAVASAVARAILPDESENEVIDLPE-----RTA 555  
DB 336 YWGHRLNLFQGAESLRIGSVSRIDAE-----SVEGHDYSAGITFTYPCMPNPTTKTS 390  
QY 556 LANRKTADYQSKKVPYLVFVASDKPRDQIGLWGSDTGTRLVTKFEHNLINRDGYOA 615  
DB 391 LIAKTEPNDTYRAKTL-----TGT---AGPAYELNDTIDTAA 424  
QY 616 GAELRL--SEKKVKYLAATKPLSHPL-----NDQRLATLGYQOEYFGHSTNGFDLST 666  
DB 425 GLEVQWADTEBAFAGKNEYLT--TSIPLEFYADRTDQAKLNPTEGFRAS----- 469  
QY 667 RTLEHEISRSIIQNGMNNRFRYSRLYRLDKLKTQAPPETWQDLPVDFVNGKPSOEALLAGV 726  
DB 470 -----IAANDSYAL----- 479  
QY 727 AVHKTADNLVNPGRYKORYSLEVGSSGLVSDANMAIARAGISGVYSGDNAYGNSRAH 786  
DB 480 -NOTFEFSSEGSITGYK-----GGAEDRLIM-----AG 507  
QY 787 QMTGGIOAGIYMSDNHNHVPRLRFAGGQDSIRGYAHDLSPISDKGYLTGGQVLAVGT 846  
DB 508 KLSGGYLVG---GSDLDIDPTTRRFRFAGGGSVGYSGYQETSPYANAGDAGTGRSVYGS 564  
QY 847 AEYNYEFPMKDLRLAVFGDIGNAVDKGFT---NDPKIAGVGVNRASPVGQVRVAVATGYK 903  
DB 565 VEARIKVTDTITGLVPFDAGVSD-GVTPDPSDIRAGAGIGLRYATPFGLRLDVAMPLE 623  
QY 904 --EGNPIKLHFFITGPF 919  
DB 624 KYDGNMFGIYAGIGOSF 641  
RESULT 17  
08UFL6  
ID 08UFL6 PRELIMINARY; PRT; 774 AA.  
AC 08UFL6;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DN Group 1 outer membrane protein.  
GN OMP1 OR ATU1381 OR AGR\_C.2554.  
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=176299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21608550; PubMed-11743193;  
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. Sr.,  
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,

Raymond C., Rouse G., Saenphimmach C., Wu Z., Romero P., Gordon D., Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan M., Perry M., Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M., Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V., Neeter E.W.,  
RT "The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58."  
RL Science 294:2317-2323(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21608551; PubMed-11743194;  
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Qurollo B., Goldman B.S., Gao Y., Askenazi M., Halling C., Mullin L., Houmel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F., William C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B., Planagan C., Crowell C., Guron J., Lomo C., Sear C., Scrub C., Cielo C., Slater S.;  
RT "Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58."  
RL Science 294:2323-2328(2001).  
DR EMBL; AE009099; AAL42387.1;  
KW EMBL; AE008064; AAK87173.1;  
SO Complete proteome.  
SQ SEQUENCE 774 AA; 84546 MW; 3C88B25F4EB537DF CRC64;

Query Match 4.9%; Score 232.5; DB 16; Length 774;  
Best Local Similarity 21.0%; Pred. No. 3e-05;  
Matches 178; Conservative 109; Mismatches 310; Indels 251; Gaps 39;

OY 165 IKRLYLARL-----NDGVNKV---PRLKAFYSSGSGSETSAIGSS 202  
DB 77 VKRLYATGVSNVSMRVSGSTLVVTVNENQLVQVFNQGNKRIKIDKLAVQTPGMPF 136  
OY 203 HQK--TEPVANIKAALEDITQESAMDLNGSIPRLROTALVAARVGYDDLSII-- 255  
DB 137 NQAVIADIAIRKIAVSAIR-----SDVEITTOIV-----SVGGGRVNIAPVINEGE 184  
OY 256 RNSIGEDVVIHDLGEVYIDYR-AVEVRGEGADKAFITVADEVPLIDVFNHGKYEY 314  
DB 185 RTKIGRIDFT---GNNSYSDGRILAIVNTKKNMISFLRK-----DIVNEDKLRA 232  
OY 315 KKLILNENASAEHGYFDGRMLDRSDVILP--NTADVSLIYDGTGYRDEVVFFITDPK 372  
DB 233 DEBALNQFYNNRGYADFRVY--SSDAVLDESKNEYTISIVDEGKRYDFGNVAVESTVP- 289  
OY 373 TNOITDPDKLPYKRELEOLTVNNGEAVNLQAVRALSNDLATRYFNVNVEIYFPER 432  
DB 290 -----GVDSSELQGLVETTRGASTSAKEVQO-SNEAISKR----- 323  
OY 433 EOIQNDVSEGOSSSRTEPAQVDESTLEPIETVELTDGILMDISPIEFSASNLIDKL 492  
DB 324 -----VAGEGYFPAVTPPRGDRDMSGMTIGYTVIVDGERAYVERIELRGTRTRDYV 376  
OY 493 NLYAAKARILYDMPDRVLAINHDDGVNRSI-----IGRISDAVASAVARATLPDE 542  
DB 377 -----IRREFD-----ISEGDAFNQTIITAAKRLLEALGYFSKYNINISAGSAPD- 421  
OY 543 SENEVIDLPERTALANRKTTPADVOYOSKKVPLVFAVSDKPRGOQIGLGMGSDTGRVLTK 602  
DB 422 -----RVYIVDVEDOSTGSFSGIGAGYSQNDGVILLEAS 454  
OY 603 FEH-NLINEDGY--QAGAELRLSEDKKGVKLYATKPLSHPLNDQLRALTVGO----- 651  
DB 455 VEKNFLNGGYIRVAAGAG--EDDARTYLSLTFEY-----FLGTRILAAGFDL 501  
OY 652 -----QEVGHSTNGEDLSRTLLEHISRSIION-----GGNNRTY 687  
DB 502 FKNOSKSDYVNYDDQGFAL-----RYTAPITENLSTTFKYTKQIYVECKGQMNNA 554  
OY 688 SLRYRLDLKLTQAPPEPTMODLPVDFVNGKPSQEBALLAGVAHKYVADNLVNPGRYORY 747  
DB 555 NL-----AEPYQAL-----IRGEDWTQISTLSMTLVNLTLDLDRM--PREGQOAL 597

OY 748 SLEVGSSGLVSDANNAIARAGISGYVSFCD--MAYGNSRAHDMGTGICAGTWSDNFHHV 805  
DB 598 TNEF--AGLGGSEYKTYAARARYTTLSDYDVLS-----LFG--QAGHVMPTGMDL 648  
OY 806 PYRLRFPAFGDQDSIRGYAHDLSPI--SDKGYLGGQVLAAGTAYN-----YEFKMD 856  
DB 649 VED-QFKFGGRQ-VRGFANDIGIPRISDS---IGCTTYFAASAEVTPAPMGVEPDF--G 701  
OY 857 LRLAVFGDIGAAYDKGFT-----NDTKAGAGVGRWASPVGQVRVAVATGKKEGCP 908  
DB 702 LRLAGFVADAGTMYGNKVSQTQVKKDMSIRASAGIGVMAASPFGPIRVDAVPIAKEGYD 761  
OY 909 IKLHFEIG 916  
DB 762 EQGFRFEG 769

RESULT 18  
O93PM2  
ID O93PM2 PRELIMINARY; PRT; 793 AA.  
AC O93PM2;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Integral outer membrane protein.  
GN D15.  
OS Haemophilus ducreyi.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;  
OC Haemophilus.  
OX NCBI\_Taxid=730;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21295095; PubMed-11401984;  
RA Thomas K.L., Leduc I., Olsen B., Thomas C.E., Cameron D.W., Elkins C.;  
RT "Cloning, Overexpression, Purification, and Immunobiology of an 85-  
kDa Outer Membrane Protein from *Haemophilus ducreyi*."  
RL Infect. Immun. 69:4438-4446(2001).  
DR EMBL; AF329831; AAK70345.1;  
DR InterPro; IPR000184; Bac\_surfAg\_D15.  
DR Pfam; PF01103; Bac\_surface\_Ag; 1.  
SQ SEQUENCE 793 AA; 88661 MW; E4E9DB62A8F9903F CRC64;

Query Match 4.9%; Score 229.5; DB 2; Length 793;  
Best Local Similarity 19.3%; Pred. No. 4.7e-05;  
Matches 173; Conservative 134; Mismatches 309; Indels 279; Gaps 44;

OY 175 DGVNK-----VRLKAFYSSGSGSETSAIGSSHQKTEPVANIKAALEDIT----- 220  
DB 28 DGQVTEFGNAILIASLVKVGQGTATDDGVYITVQKLSQNRFFENVASARSGQTLVIVAE 87  
OY 221 ---QESAMDLNGSIPR-----LROTALVAA-----RAVGYDIDL 252  
DB 88 PLINNLITIKGNNAIPKNALEONKANLIVAGEYDKAKLEAFQALVDHYHTMGRYQADI 147  
OY 253 SIIR--NSIGEDVYIH-DLGEVYIDYRAVEYRGEGADKAFITVADEVPLIIG---D 305  
DB 148 QITTPPNNNGSINVELNITEGELAYVK---KINFGNNAFSDDELKELEIKPNAPMWN 203  
OY 306 VFHNGKYETKK-----NLLENASAEHGYFDGRMLDRSDVUILPNDVADSLY--DTGQ 358  
DB 204 IFESSKFOQDEINKDELILRDYMDHG--AKFTLADTVKFNENKTEVDLTYKKINEGSO 261  
OY 359 YRFDEVVFEITDPKTNQLTTPDKLPYKRELEOLTVNMGGEAVNLQAVRALSNDLIATR 418  
DB 262 YNISEK-----RIIGDTQKLDNE--LNQLLT-----HFKAQO 291  
OY 419 YFNMYNTELVFEREQIONDOVSEGOSSSRTEPAQVDESTLEPIETVELTDGILMDIS 478  
DB 292 LFRKTELSTIEIOIKOIIIGDR---GYGSAK-----VDLY 322  
OY 479 PIEFSASNLIDKLMLVAAKARHLVDMDDPDRVLAINHDDGVNKRSLTGR-----ISD 529  
DB 323 KFNEDHIVQ--INFIYDAGRIT--VKKIREGNDVADSTLIRREMQOEGAMLST 376

```

Query Match      4 8%:  Score 225;  DB 16;  Length 861;
Best Local Similarity 19.7%:  Pred. No. 0.0001;
Matches 189;  Conservative 130;  Mismatches 393;  Indels 242;  Gaps 38;

QY  93 SPINRIGSPPGLGDMVETETPLS-TEELFADESTFMGJNPNDIPIYEGQEDPNSEV 151
      || : : : : : : : : : : : : : : : : : : : : : : : : : :
      11 SPYVLLLTSLGVGASPAQATOFSPILNMLADNDNFTLVSPSSSL-NEFGGPPSRFA 60

```

QY	152	VPTPLLEPKELLKRLRLARLENDGVANKVPRYLAKAFYSSSGEFLSAGSHQKTEPYAN	211
Db	70	SNVSTFLQEAAPSVINALNGE--NGEISVIPR-ETEQIDGRIGQTMETISA-----GN	118
QY	212	IKAALEDITQESAMQ-----LNGSIPRLRQALVAARAVGYDDIDLSIRNSIGEVNII	266
Db	119	LQVGYDDLPAPAPVUSAEALQANLEPHGNNAVAQVAPRAQVQEDNLAETKTDNQTDBV-	177
QY	267	HDLGER-----VYIDYRAVEYRGEGADKAFETTVADVEPPLLIGDV-----FHNGK	311
Db	178	----EPOSPILMAQAQVAEEBEVAEEAEKEATETTGATVEETPEPTPTPADPPTNDETPGR	233
QY	312	YETKKNLIEMNSAEHGYFDGRMLRSVDVYIIPDNTAD-----VSLYDTGYGRF	361
Db	234	TQTLSPSTFPAS-----PSFTTPAPAEERPRVLVSEVLVYGTPEL	274
QY	362	DEVPETPIDPKTNOLTTPDPKLPYKRELLBOLLVYNNGEANLQAVRALSNDLITATYFN	422
Db	275	ELVAV-----NAIRTPGKRTTQTQLOEDV-----NAIATATYFS	309
QY	422	MYNTEIVPEREQIINDQVSEFOSSSS-----RTEPAQ-----VDESTLE-	466
Db	310	NVR---VAPSDPPL-GVAVTEFEVQANPYFTGLNIRTYETAEKGERILLPQEVNDETFGEQ	365
QY	462	--PVIEYVELDGTIIMDISPIEFSASNLIDQKLNLVAKARHLYDMEDRYLAINHGGV	519
Db	366	YKGIINLRELEQ-----IKTINENYSGQDYLAQVQSP-----QVAGDQGVTLVIAEGI	416
QY	520	NNSILGRISDAVSAVARAILPDESENEVIDLPERTALANKR---TPADVYQSKV-----	571
Db	417	VENIOVREFD-----SEDEVEQGTSDQFITTREMRLLKPDGVNRRNNAQTDLQ	463
QY	572	-----PLYEFAVSD--KPRDGOIGLGWG--SPMGTRLYTKFPHNL	607
Db	464	RYSISGLFEDYRLSPNBSDPTEYIVANDVYEGNTGSLAAGGSISSSGLPGLTISYERN	523
QY	608	INRDGYQAGAEILRLSEDKKGVKLYATKPL--SHPLNDQLRATLGYQOEVEGHSTNGFDLS	665
Db	524	LGNNQOTIGVBAQVQORELLPQVSTFDPMWIGDQPFRTSYTANL-FRRRTISLVEFDGDS	582
QY	666	TRTLEHEISRIQNG-----GWNRTYSLRTRBLDKLKTQAPPEWMDL	708
Db	583	IRTFNGFSDPVRVETGLTFEPRIDADVAPRPMWRSLAGFGYONVRIENAAAGALSFFSA	642
QY	709	PVDFWNGKPR-----SQEALLAGVAVHAKYADLVLPMPMGYQORYSLLEVGGSGIVSDAN	761
Db	643	PLNGFNPSPLFSFDYGVDELFTLSFGSASQDNRRNNAALDPTGSLVRFGEA--QTPVGTGN	700
QY	762	MAIARAGISGVYSF-----GDNAVGSNRAHMQTGGIOAGYIWSDFNHNVPKLRF	811
Db	701	IMMTR--LRGSYTYIIPNMMLDLGLFGLVSTGQQTQYAFANVQAGTVLGD---LPRYLEAF	754
QY	812	FAGDQDSIRGYAHDSLSPISDKGYLTGGOVLAVGTAEYNEFMKDLRLAVFGDIGN---	867
Db	755	ILGGSNSVYRGY-----QEGELGNGRSFFPQATAYEPRPIIAAVGALFVYDGSNLGS	805
QY	868	-----AYDKGFTNDKIKGAGGVGVMASPVQGVARNDAVATYKREGBNPKLHFFGTGP	919
Db	806	QGAVPGRPAIVKGLPG--SGVGyGLGVRIQSPVGRIRDL--GFTGGEES-RINFGIGEKRF	861
RESULT 20			
Q8YHHO			
ID	Q8YHHO	PRELIMINARY;	PRT; 781 AA.
AC	Q8YHHO;		
DT	01-MAR-2002 (Tremblrel, 20, Created)		
DT	01-MAR-2002 (Tremblrel, 20, last sequence update)		
DT	01-JUN-2002 (Tremblrel, 21, last annotation update)		
DE	Outer membrane protein.		
GN	BME10830.		
OS	Brucella melitensis.		
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;		
OC	Brucellaceae; Brucella.		

OX NCBI\_TaxID=29459;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=16M / ATCC 23456 / Biotype 1;  
 RX MEDLINE=20020109; PubMed=11756688;  
 RA Delvecchio V.G., Kaputal V., Redkar R.J., Patra G., Muijer C., Los T.,  
 Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,  
 Jablonksi L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,  
 Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Leesson J.-J.,  
 RA Haselkorn R., Kyriplides N., Overbeek R.,  
 RT "The genome sequence of the facultative intracellular pathogen  
 Brucella melitensis";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
 DR EMBL: AE009524; AAL52011.1;  
 DR InterPro: IPR000184; Bac\_surfAg\_D15.  
 DR Pfam: PF01103; Bac\_surface\_Ag:1.  
 KM Complete proteome.  
 SO SEQUENCE 781 AA; 85918 MW; E3C7385CCB1DB3A CRC64;

Query Match 4.7%; Score 223.5; DB 16; Length 781;  
 Best Local Similarity 20.9%; Pred. No. 0.0001;  
 Matches 180; Conservative 116; Mismatches 302; Indels 263; Gaps 45;

OY 165 IKRLIA-----RLFNDGVNKKVRLAK-----FYOSSGSETSAIGSSHQKTEPYAN 211  
 DB 78 VKRLFAAGLPSDVIRHSGSTFLIVQKERSVNNVLFQGNKKIKOPDLARAQV-LKPRAP 136  
 OY 212 IKALEDITQESAMDLNGSIPRLQTLVAARAVGYDIDLSIRNSIGEVDVTHDGE 271  
 DB 137 FDMATMAOKAEIAKAVSHIG--RSDATVNAKRV-----DL--GQGRNV--VYEINE 183  
 OY 272 PYVIDRAVEVEGADAKFT--TVADEVP-----LLIGDVFHNGKYEKKNLIE 320  
 DB 184 GSRTKIANIEFVG-----NQAFSGRLRDVISTKSNPLMLTRNDVYDEGRLOADETLR 239  
 OY 321 NASEHEYFPGRMIDRSVDVILPNTADVSLITDTGYQVREDEY-VFTIDPKTNQLTTD 379  
 DB 240 RFTYNNRGYAFRLSSNAVLDPSTNEYTITITVDEGPRYFGDVSVESTVD----- 290  
 OY 380 PDKLPVARELLEQLLVNMGAEVNLQAVRLANDLITRFNMVNTIIVEPEREQIONDQ 439  
 DB 291 ----GVDTQALDLVLTTRTKPKISAKIE-----DSVLSTVESVAGSG 329  
 OY 440 VSEFSSSSSTEPAPQVDESTLEPIVETVELTD-GILMDISPIEFSANLQDKLNLVAAK 498  
 DB 330 YAF-----AKVEP-RGDRNFENHTISVYVSDGPRAYIORIEIRGDKTRDYV----- 377  
 OY 499 ARHLVMPDRVLAINHDDGVNSIIGR-----ISDAVASAVARAILPDESENEVI---D 549  
 DB 378 IRREFD-----LNEGDAFNQVAVQRAKRLLEVLDFEQTVNISTAPGSEPPQVILVVD 429  
 OY 550 LPERITALNRKTPADVYQSKVPLVYFVASDKRDQOIGIGMKSMDG-----TRLVT 601  
 DB 430 VVEKST-----GEFSIGGGYTTGGESPGAQVEAALIT 460  
 OY 602 KEENHLINRGY--QAGAEFLRLSEDKKVKLYATKPL-----SHPLND- 642  
 DB 461 --ERNFLGRCQYIRISAGAG---QDDMRNVGLSFTEPYGLRSLAGFDVFRSYVND 515  
 OY 643 ----QLRAVLGIQOEVEFGSTNG-----FDLSTRLEHEISIT---ONGSMNR- 685  
 DB 516 YDEVTGCTIRGLPTTDFNSAGIAYSLVQEKYDLEFRGDAENVYAPALLEAENSFWLS 575  
 OY 686 --TYSLRY-RLDKLKTQAPPETMODLPVDFVNGKPSQE-ALLAGVA--VHKIVADNLVMP 739  
 DB 576 SVSYSLTITSSIDDIKN-----PHDGLGKFIQEPFAGIGDAKYKTKTFKGN--- 621  
 OY 740 MCGYRQRYSLVSGSLVSDANMAIARAGISGVSYFGSDAAYGSNRHQAHTGTGTOAGIWS 799  
 DB 622 ---YVOTLSQEADIVGLG-----VGAGYIHFEFGD-----GVRI----- 653  
 OY 800 DNFNHNVPYKRLRFAGGDSIRGYAHDSLSPISD-----KGYLTGGQVLAVGTAEVNEEF-- 853

DB 654 -----FDLFKNSDILRGFKFNGIGIPIYODAKNKGRIYMGTTYSGAIEVOFPMPV 704  
 OY 854 ---MKDLRAVFGDIGNAVYDKGFTNDT-----KIGAGVGV--RMASPVQGVENVAT 900  
 DB 705 LPESLGVGAFFADAAATLYG-----NDTPDISGDDKKLRASVGSVLMMSPFQPLRFDTAF 760  
 OY 901 GY-KEEGNPIK-LHFFIGTTPF 919  
 DB 761 PVAKADTDKQVNFNGVSTKF 781

# RESULT 21

ID Q98MC3 PRELIMINARY; PRT; 794 AA.  
 AC Q98MC3;  
 DT 01-OCT-2001 (TREMBLrel. 18, Created)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE Outer membrane protein.  
 GN ML0636.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_Taxid=381;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=MAF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Matsubae A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsuno M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,  
 RA Takeuchi C., Tamada Y., Tabata S.,  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 Mesorhizobium loti";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AP002995; BAB48190.1;  
 DR InterPro: IPR000184; Bac\_surfAg\_D15.  
 DR Pfam: PF01103; Bac\_surface\_Ag:1.  
 KM Complete proteome.  
 SO SEQUENCE 794 AA; 85741 MW; BABE4A1BC8FEEDC6 CRC64;

Query Match 4.7%; Score 223.5; DB 16; Length 794;  
 Best Local Similarity 20.4%; Pred. No. 0.00011;  
 Matches 177; Conservative 130; Mismatches 321; Indels 241; Gaps 44;

OY 156 TLPEPK-----PGLTKRLYAR-LFNDGVNKKVRLAKAFYOSSGSETSAIGSSHQKTE 207  
 DB 62 TIKPKAFSSSDVDAVAKALFGTGLFSD-----VQINQVGSITLVVVAEKVV 109  
 OY 208 PYANIK--ALEDITQESAMDLNGSIPR--LRQTL-----VAVARAVGYDIDLSII 255  
 DB 110 NQVLFQGNKKLKNALAAVQLK---PRGTFSGQATLDSDEVAVKAAYRRIGRDAGVTTQ 166  
 OY 256 RNSIGE--VDVLIHHDGEPIYIDYRAVEVREGA-----DKAFTTVADEVPLTI-GDV 306  
 DB 167 VVELGDNRVNVVFN--TTEGDRQIQAIAINFGVNSAYSSRLSDVINTRKSSWVSFIIRDYV 225  
 OY 307 FHHGKETKKNLLENASAEHGYFDGRMLDRSVVDVILPNTADVSLYDGTGYQVREDEYV 366  
 DB 226 YDEDKLRADQELLRRYYNHNHGAADPQVYSAVGEIDNATKYYITVTVQGERINFQDISV 285  
 OY 367 FTIDPKTNQLTTDPDKLPVARELLEQLLVNMGAEVNLQAVR---ALSNDLIATRY-FN 421  
 DB 286 ESTIRE-----VDSKSLSESVETHHKGDVYNAKADVEDSITALLEKVAAGSYATA 333  
 OY 422 MNTETVPEREQIONDQVSFEQSSSSTEPAPQVDESTLEPIVETVELTGLIMDISPIE 481  
 DB 334 QVT-----PRGR-----NFENHTIS-----VVYITVD--QGTKAVIERIE 366  
 OY 482 FASNLIDQKLNLAAKARHLVMDPDRVLAINHDDGVNRSITLGRISDAVSAVAAILPD 541  
 DB 367 IRGNDTRDYV-----IRREFD-----VSEGDAFNQVLIOR-----AKRLED 404

```

QY 542 ESENEVIDLPRTALANKRTPADVYQSKRVPYVVASDKPRDQIGLWGSNCTRLVT 601
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 405 LNYFDKVEI-----STVPGS--QPDQVVLVYDVVEKSTGEFSVGAGY--STGDSAGP 453
QY 602 KFEHNLNRDGYQAGAEIRLS-----EDKKGVKLATKPLSLPNDQLATIGLQOEVGCH 657
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 454 SVESSITERNLGRGQPKLSAGGKNSRDYSVSFTPE--YFLGRRIA-----500
QY 658 STNGFDLSTRLLEH-----ISRSITONGGMMNRTYSLARYLCLKTQA 700
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 501 ---GFDLYKSTRRENNNDSDTVGATVAFGLPITNSITITOLATNISOE--KIKVD---NSC 553
QY 701 PPEWTQDLPVDFVNGKPSQEAALLAGVA-----VHKTVADNLVPMRCYRQRY 747
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 554 GPTDPNPGTCN---ISQAILDGIAESPWKSSVSLGLVYNTI--DMKNPHEGITY---606
QY 748 SLEVGSSGLVSDAMMAIARAGISGVSF-----GDMNYSNNAHQW-----TGCIQAGYI 797
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 607 -----ANTTVEVAGVGDAKFKITGRGSIYQTLSEOLDVLGLISGG--AGHV 652
QY 798 ---WSDN---FNHVPYRLRFAGGDSIRGYAHDLSLPSID--KGYLTGGOVLAAGTAE 848
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 653 EGYNDQDLRFEDH-----FQSTDMIRKGFATGIGPVASGTSGLDGLGTTTFMASAE 705
QY 849 YNVEF-----MKDLRAVFGDIGNAYDKGFTNDT-----KIGAGYGVWASFPVG 892
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 706 AQFPLVPYPSFGLRGAVFADATLYGKIKADQTLVDDSTGMKLRASVGVGLMMASFPFG 765
QY 893 QVRDVATGKVEGNNP--IKLHFTIGTPE 919
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 766 PIRIDYAIPIVAKKASDDVOEFNFISTRE 794

```

## RESULT 22

```

O9JUM5 PRELIMINARY: PRT: 615 AA.
AC O9JUM5;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Putative outer membrane protein.
GN NMA0296.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parikhil J., Achtmann M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
DR EMBL; AL162752; CAB83602.1; -.
KW Complete proteome.
SQ SEQUENCE 615 AA; 67775 MW; B9A64607638E1C6B CRC64;

```

Query Match 4.7%; Score 220; DB 16; Length 615;

Best Local Similarity 22.3%; Pred. No. 0.00011;

Matches 11; Conservative 72; Mismatches 236; Indels 148; Gaps 19;

```

QY 462 PVIEVELTDLGILMDISPIEFSASNLIDQKILNTVAARAKHLVMPDRVLAINHDDGYNR 521
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 48 PDRESVYLRKPKFPRIDTQSEIKDWVEHPLITQOEVLDEQGFGLAEAPDNVKT 107
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 522 ST--LGRISAVSA-----VARAILPDESENEVIDLPRTALAN 558
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 108 MLRSKGYESSKRVSLTEKDQAVTHTTGPRTKIANGVAILGLDILSDGNLAERYNALEN 167
QY 559 RRTPA-----DVYOSKR-----VPLVVFASDPK 582
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 168 MQQPVGSDPDQDSWENSKTSYLGAVTKRAYPLAKIGNRAVNPPTATADLNVVYDSGRP 227
QY 563 RDQIGLWGSDDTGR-----LYTKFEHNLNRDGYQAGAEI- 619
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 228 ---LAFGDFEITGTRPEQIVSGLARFQPTPYDLDLLDFQOAL--EQNGHYSGASVQ 282
QY 620 ---RLSEDKKGVKLYAKRPLSLPNDQLAT-----LG-----YQDEVFCHSTNG 661
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 283 ADFRLQGDVRPVKVSYEVKRRHKLETGIRLDSXGLGKIAYDYNNLFNKGYISGVYWD 342
QY 662 FDLSTRLLEHRSITONGGW---NRTYSILRYLCLKTQAPEWTQDLPVDFVNGKPS 718
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 343 MDKYETTLAAGISQPRNRCNWTNSVSYNSTTONLEKRAFSGGIWVRDRAGIDARLG 402
QY 719 QEALLAGVAV-----HKT-----VADNLVPMRCYRQRYSLVEGSSGLVSD 759
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 AEFLAEGRKIKGSDIDLGNSHATMLTASMKRQLANVLHPENGHYLDGKIGTTLGAFISS 462
QY 760 ANM--AIRAGISGVSYFGDNAYGSNRAHQMTGCI---QAGY--IWSDNFNHVPYRLRPF 812
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 463 TALIRTSARAG-----YFETPENKRLGTFPIIRGOAGYTVARDNAN--VPSGLMFR 510
QY 813 AGGDSIRGYAHDLSLPSIDKGYLTGGOVLAAGTAEVYEFMKDLRAVFGDIGNAYDKG 872
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 511 SGGASSVRYGELDSTIGLAGPNSVLPERRALLVGSLEYDLPFTRLTSGAVFHDMDGAANF 570
QY 873 FTNDIKIGAGYGVWASFPVGQVRVDVATGVKKEGNNPILHFTIGTPE 919
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 571 KRMKLKHSGLGVWRWFSPLAPFSFDIAYGHSDK--KIRMHISLGTRF 615

```

## RESULT 23

```

O9BK30 PRELIMINARY: PRT: 617 AA.
AC O9BK30;
DT 01-OCT-2001 (TREMblrel. 18, Created)
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical protein ml1662.
GN ML1662.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002998; BAB48984.1; -.
DR Interpro: IPR000184; Bac_surfact_D15.
DR Interpro: IPR003975; Shal_channel.
DR Pfam: PF01103; Bac_surface_Ag_1.
DR PRINTS: PR01497; SHALCHANNEL.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 617 AA; 65759 MW; 83FB8C176201A444 CRC64;

```

Query Match 4.7%; Score 220; DB 16; Length 617;

Best Local Similarity 22.2%; Pred. No. 0.00011;

Matches 150; Conservative 95; Mismatches 246; Indels 186; Gaps 30;

```

QY 313 ETKKNLI---ENASAEHGYFDGRWLDRSVIILPDNTADVSLIYDTGYGRFDEVVFETI 369

```

```

Db      49 DADKDLVRLKLENASALKGGED--RPVSGSLGLMAKARSDEQLVAALFADAREGVTTIT 107
      370 DPK-TNOLTTD-----PKLPVKRELBELLTVNNGEALNLOAVRLSND---LIATRF 420
      108 DCKPLDELPPAEPFGPOPIV-----TINIAVGPKFTLTIR-LEDDAAGLMSADVG 159
      421 MNVNT-----IVPPERQ-----IONQVSFEQSSSSRTPEPAVDSTLEP 462
      160 LISGDACSGAVALKAELIVRTLKEQGRPLAKVTDRRIADATSTLDVTLTV---AAGP 216
      463 V-----IEVELTDGILMDI-----SPIEFSASNLIDQKLNLVAKARHLYDM 505
      217 VAGGATTVTEGKEVDKEDFTETEMTGLKRGKQSPQEISDAR----- 257
      506 PDDRVLAINHDDGVNRSILGRISDAVSAVAAILPDESENEVIDLPERTALANKRTPADV 565
      258 --DRLLALE---VFNSYTEREADKLDA----- 279
      566 YQSKKVPPLVVFASDPKPRDGOIGLGMGSDTGTRLVTKFEHNLINRDGYQAGAEIRLSEDK 625
      280 --DGNIPIGVQVSEKRPYFGLGTFSTBEGLLEGYMGH---RNLFGHAEKLRIDGAI 333
      626 KGVLYLTKPLSHPLNDOLRATLTGYOE---VFHSTNGFDLSTRTLEHETSRIIQNGG 682
      334 SGIG-----SNNLSD-LANYNAGIMEKPGVIGPASKFFAGFKTVLEHPADYHFSYKG 385
      683 WNRYSRLRYRLDKLTKTOAPETMODLPVDYFN-----CKPSOEALLAGVAVHKTIV--ADN 735
      386 ---STGLSTELDKQQTYS-----AEVALDYSRITDAFEKHTY--LIASVPLQYVYDNRDS 435
      736 LVNPMRGYR-----QRYSLSEVGSSGLVSDANMAIARAGISGVYSEGDNAVSGNRHOMT 789
      436 RLNPTRBGRFLVLAEPYDIMGSAFL-----KLKGESATVQSIDTAKFVLAER-- 485
      790 GGIGAGIWMDFNHHVYRLRFEPAGDOSITRGVANDLSPLISDKGYLTGGVLAAGTAEY 849
      486 --VAIGIVTGTGLQNVADRFRFYSGGGSVGYVQVIGIRPFGOPITGLSF----- 536
      850 NYEFMKDLRLAV-----FGDIGNAVYDKGFTN--DTKIGAGVGVMMASVGVQVRDVA 899
      537 -FETSVMRIAVYDTIGIVFPVADGIVSTKSVNFSVDKVGAGVLYRVPFGLRIDAA 595
      900 TGVKEEGNPIKLHFFIG 916
      596 VPLNRDPNDPHFGIYAG 612

```

RESULT 24

092LN9 PRELIMINARY; PRT; 618 AA.

AC 092LN9; 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Hypothetical protein R02996.

GN R02996 OR SMC03097.

OS Rhizobium meliloti (Sinorhizobium meliloti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Sinorhizobium.

NCBI\_TaxId=382;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=21396507; PubMed=11481430;

RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,

RA Godle T., Goffeau A., Kahn D., Kiss E., Leleau V., Masuy D.,

RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsberger U.,

RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galbetti F.;

RT "Analysis of the chromosome sequence of the legume symbiont

Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001)."

```

DR EMBL: AL591792; CAC47575.1; -
DR InterPro: IPR000184; Bac_surfAg_D15.
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR001412; trna-synL-1.
DR Pfam: PF01103; Bac_surfAg_1.
DR PROSITE: PS00178; AA-trnaLIGASE-T; UNKNOWN_1.
DR PROSITE: PS00225; CRYSTALLIN-BETAGAMMA; UNKNOWN_1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 618 AA; 65656 MW; E51276B39B1FB78 CRC64;

Query Match      4.7%; Score 220; DB 16; Length 618;
Best Local Similarity 19.6%; Pred. No. 0.000117;
Matches 168; Conservative 108; Mismatches 279; Indels 304; Gaps 35;

      117 PLISLE-----LF-----AOESTEMGINPNDY-----PEYGEOPNSSEVVP 155
      8 PVSVEQAHAFKIFGRPFESAEEVQV-IDPRYITLTFEPGTDEBLEALENSQLVQD 66
      156 TLEPKPGLIKRLVRLKNDGVNKKVRLKAKFYOSSGSETSAIGSSHQKTEPVANIKAA 215
      67 QEEPVSGDLGLAIKARDRD-----RLAVLYEKARVGGTVSILVNGD----- 110
      216 LEDITQESAMDLNGSIPLRLROTALVAARAVGYDDLSITINSIGEVDYIHDGEPIYI 275
      111 IDSLPPDAFPDGOVPVAVVYARGPATLTGTVLE-----GDAARL 152
      276 DYRAVEV-RGEGADKAPATTVADEVPLIGDVFNHGKYETKKNLIENASAEHGFDRWL 334
      153 DPAAYDLKRGARADSTLIIKAGEQI--VNDL-----KEGSRPLAKLA----- 192
      335 DRSDVILPDMTAVVSLIYDTGTQYREDEVVEFTIDPKTN-----QLTDPDKLPVREL 389
      193 ERSVADHATSTVYVTLRADGG-----PVAPVGNLTVSGARTVDPD-----F 234
      390 LEQLLTVNMGAVNLOAVRLSNDLIATRYFNMYNTELVPPERQIQNDQVSFEQSSSR 449
      235 VKDYSRLNHGFRYPENIRKAEERLQNLNFFSV----- 268
      450 TEPAQVDESTLEPIETVELTDGILMDISPIEFSASNLIDQKLNLVAKARHLYDMPDDR 509
      269 -----TINEADGLAPDGT-----IPMNIQVSEGNHRFFGGCQ 301
      510 VLAINHDDV-----NRSLIGRISDAVSAVARAILPDESENEVIDLPERTALANKRTP 562
      302 V---STTDGLGLOGYMGHRLFGRAESL-----RIEGSVDRIGETTIVAGLDYS 347
      563 ADVYOSKK---VPLVYVFASDPKPRDGOIGLGMGSDTGTRLVTKFEHNLINRDGY----- 614
      348 AGILFAKFGAFGPASTFTASVKA-----IYDPDAVSAKTVT 384
      615 --AGAEIRLS-EDKKGKLYATKPLSHPLNDOLRATLTGYO--OEVFHSNNGFDLSTRTL 669
      385 AAGAAFFELSPEDTFSV-----GAEVGMADVDDARG----- 415
      670 EHIISIIIONGWNKRTYSLRYRLDKLTKTOAPPETMODLPVDVNGKPSQEALLAGAVAH 729
      416 ---SNSYI-----TAA-----LPFEVY----- 429
      730 KTVADNLVMPMRGYR-----ORYSLEVSGSLVSDANMALIARAGISGVYSEGDNAVSGNSRA 785
      430 RQARDKLNPTGCRYRALINAKPSYEIEGKTFSS-----FEASAGYV-----AFGTEKR 479
      786 HONTGGIQAG-YIWSDNFNHHVYRLRFEPAGDOSIRGVAHDSLPSIDSKGYLTGGVLA 844
      480 FVLAKGLAGVILVGDDELSDIPATRTREFLGGGSVNGYVQIETSPRDADDELTGGSYYS 539
      845 GAAEYVEEMKDLRLAVPDIGNADKGT--NDTKIGAGVGVMMASVGVQVRVADATGV 902
      540 GSELEARIATVDTIGVVPFIDAQTVSDTAPDSDIRAGAGIGIRVATPGPIRLDPAVL 599
      903 KE--EGNPIKLHFFIGTGP 919
      600 NKYPGSTDYGIYAGIGQSF 618

```

```

RESULT 25
ID 09P128 PRELIMINARY: PRT: 739 AA.
AC 09P128:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Outer membrane protein.
GN CJO129C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCCT 11168.
RX MEDLINE=20150912; PubMed=10688204;
RA Parhill J., Wren B.W., Mungall K., Kelley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagsals K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajadream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett J.B.G.;
RA "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
DR EMBL: AL139074; CAB72613.1;
DR InterPro: IPR000184; Bac_surfag_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 739 AA; 83197 MW; B3C1891A8AA781BF CRC64;

Query Match 4.6%; Score 216.5; DB 16; Length 739;
Best Local Similarity 18.1%; Pred. No. 0.00024;
Matches 172; Conservative 142; Mismatches 317; Indels 317; Gaps 43;

QY 68 ARLNA-----GLNAKPSQALDVNFDOSPIRIGESPPGLDMVIEETTPLS 119
DB 13 AMANATIKIDIKFIGNHLSNTSAINIAGL-----KIGEEINPAKIMTALIN---LY 61
QY 120 LELFAOESTEMGINDPIPEYQGPQNP-SEYVVPPILEPEKPL-----IKRLYARLF 173
DB 62 KONTENTNAVE---NNGGLEITIVTEKPIAVYTTIGIASNDRKQVESILGKR--GTL 116
QY 174 NDGVNK--VPRLKAFYQSS-----OSGESAIGSSH--- 203
DB 117 DEGNKKAERIKAYEAKSYFDITVEYKKKLTLENTDGELEFYVRNGENIITIDNVHLSG 176
QY 204 OKTEPYANIKAALEDITQF-----SANDLNGSIPRLROTALVAARAGYDI 250
DB 177 AKKFSYSIEPAVYVNRKEFEFGMMWRNDGKLKVFELSDSSRIADEVM---KKYLDV 232
QY 251 DLS-----IIRNSIGEDVVIHIDGEPYIDYRAVEVEGADKAFPTTVADENPLIG 304
DB 233 QVSSPYLKTYYDTDYANLYTFKE--GKPKYI--KSIISINPLFDDKQNAQTVKDLSSAG 289
QY 305 DVFNHGKYEYTKNLIENASAEHGYFDGRWLDRSVDYILPDNTADVSLIYDTGYRFDYV 364
DB 290 KTINIEDIRKDKKTIETQSADIGY---AFVEYYPDIQKKDQTOEATVFKV---IPHDV 343
QY 365 VFTIDPTNLTTPDKLIPVKRELLEDLLITVNGEAVYMLQAVRLSNDLIATRYFNMYN 424
DB 344 YIRNVLISGNSRTVD---RVIRRELY-----ITEGNLYNRDTLSSKNALKTKTSYFDVN 395
QY 425 TEIYPEREQIQNDQVSPFQSSSRTPEAQVDESLPIETIEVEL-----TD 471
DB 396 -----IKEEKYDDTHIDLVLVDKEASTGAISGGIGYSSD 430
QY 472 GILMDISPIEFSASNLIDKLNLVAKAKAHLVYDMPDRLVLAINHDDGVNRSITLGRISDAV 531
DB 431 GILNLS---LSDFNIFOSGI-----KSSVSVYDKSD-----TLSGRIS--- 466

```

```

QY 532 SAVARAILPDESENEVIDLPERTALANKRTPADVQSKKVPYVVASDKPR--DGI 589
DB 467 -----LVN-----PRVDSQYSL 479
QY 590 GWSQDGTGLVYTFEHLINRDKGYQAGAEILRLEDKKGKVLKATPLSHPLNDQLRATLG 649
DB 480 G-----GT-----LYS-----ND----- 487
QY 650 YQOGEVCHSTNGFDLSTRTLEHESRSIIQNGGNNRTSYRYRLDKLKTQAPETWQDL 709
DB 488 YEMDNVSEKNYGFPI---TIGQPARLY-----NVSILYNEQ----- 522
QY 710 VDFVNGKPSQEALLAGYAVHKTAVADNLVNPARGYRQRYSLSEVGSGIVDANNAIARAGI 769
DB 523 SDIYHLSP--LRTGYELQKSIKSS--ITPAITFNDDVYLRPSGII--ASISLEAGL 577
QY 770 SGYVSFGDNAYGSRNRAHQTGTIO--AGY-----INSDNENHVPYRLRFPAGG 815
DB 578 GGDQEF-----ISSSKFNFYQGLQDYIGDLYIRKASFYKVDQGY--LPINQRYLGG 631
QY 816 DOSIRGAHDSLPISDKGYLTGGQVLAAGTAENYEFMDLR--AVFEDIGNAYKGF 873
DB 632 IKSIRGESKTVSPKQMGDEIGTTAFANSVELSPFLDIRIKLGRSVFEDYGMIGRKNL 691
QY 874 TNDTKIGAGVVRWASPVGOVRVDVATGV--KEGNPIKLFHFTGPF 919
DB 692 DEIKRSTGIGIEMITIPGLOLVFPAKPLNDKKGDDTINSEFNLGTRF 739

```

```

RESULT 26
ID 09A7W6 PRELIMINARY: PRT: 628 AA.
AC 09A7W6:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein CC1603.
GN CC1603.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15.
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uitterback T., Tran K., Wolf A., Yamatheva J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
DL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005835; AAK23582.1;
DR TIGR: CC1603;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 628 AA; 68203 MW; 92DAD1722C6F82295 CRC64;

```

```

Query Match 4.6%; Score 216; DB 16; Length 628;
Best Local Similarity 21.8%; Pred. No. 0.0002;
Matches 110; Conservative 87; Mismatches 236; Indels 72; Gaps 17;

QY 456 DESTLEPIVETVELDGLMDISPIEFSASNLIDKLN-----VAAKARHLY-----D 504
DB 155 DEGVQRQAVAAARLTEGEGRSADYVAGGRIVAGYADVAAEPREYVVDHADRTY 214
QY 505 MPDDRLA-----INHDDGV-----NRSLIGRISDAVSAVARAILPDESENEVIDLPERT 554
DB 215 RPTFRIMGELVRLGVVDVYTKGRTPNEVWGRILAPVVA-----GGVYPPEDVA 262
QY 555 ALANRKTADVQSKKV-----PLYVFAVDKPRDQIGLQWSQDGTGLV 600

```

```

Db 263 ELERRLRDTAVYDSISVSLAGTDKASAEGRYPVVVTLSDRARFTIELGAGSTSEGCACVD 322
Oy 601 TKFEHNLININDGQVAGAELELSEDKKGVKLYATKPLSHPLNDOLRATLGYQOEVFGHSTN 660
Db 323 ARWR--YNNQKRAADTTTVALRFAKLBORLGAELSLPHMWRPQ--QTLKLNSSVFRNDTD 378
Oy 661 GFDLSTRTLEHISRSITIQNGMNRRTSLRYRLDKLTKTOAPPEWMDLPDVFNGKPSQE 720
Db 379 AYNETGATVGVDLTR--RQTTAVRTTCGVSPFSDLSQTKEQVNRN-----GLIAGRKLNL 429
Oy 721 ALLAGVAVH--TVADNLVPMRGYRORYSLEVSSGLVDANNAIAR-AGISGVSYSGDN 778
Db 430 ATLAGLAAYAMFSDSDLLDPRKRCGRMLETRA--PTYVAGDTSVPLYKLACGGSAYL---- 483
Oy 779 AYGSNRAHOMGTGQIAGYIYSDNPNHVPYRLRFAGDQSIKGYADSLSPISDKGYLTG 838
Db 484 PFQKQDSTVLAARVKALGAILGAGLLDVPASRRFSGGSGVSGYAYQALIGPRLSDNTPQ 543
Oy 839 GQVLAAGTAEYNEFMKDLRLAVFGDIG--NAVDCGFTNTKIGAGYGVWMAAPVQGVY 896
Db 544 GISLVETSEFVRQKIDRMGSAFVDAAGALGTHETPQREDFRAGAGLGVATDLGFGPIRA 603
Oy 897 DVATGV-KEGAP-IKLHFFIGTPE 919
Db 604 DIAAPLGRKCDPKFOIYLSIGSF 628

```

## RESULT 27

09JXB7

PRELIMINARY: PRT: 635 AA.

```

AC 09JXB7
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Hypothetical protein NMB2134.
GN NMB2134.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson M.C., Gunn M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappunli R., Venter J.C.,
RA *Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.
RL Science 287:1809-1815(2000).
DR EMBL: AE002361; AAF42442.1;
DR TIGR: NMB2134;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 635 AA: 69801 MW: 28371.647A9DA98 CRC64;

```

Query Match 4.58; Score 215; DB 16; Length 635;

Best Local Similarity 22.18; Pred. No. 0.00023;

Matches 132; Conservative 73; Mismatches 225; Indels 166; Gaps 20;

```

Oy 462 PVITVELTIOIIMDISPIESASNLQDKNLVAAARHLIYDMPDRVLAINHDDGVNR 521
Db 68 PDETSVALKRPFPVLIDTQDSEIKDWEHMLPLITQOEEVLDEQTFLEAEAPDNVKT 127
Oy 522 SI--LGRISDAVSA-----VARAILPDSENEVIDLPERITALAN 558
Db 128 MLRSKGVFSSKSVSLTEKDAVTVHTTGPRTKIANVGAAILGLDLSGNTLAEYVRNLN 187
Oy 559 RKTPA-----DYOQSK-----VPLYVFAVSDKP 582

```

```

Db 188 WQQPVGSDFDQDQSWENSKTSVLGAVTRKAYPLAKIGNTOAANPDATADLVNVDGSRP 247
Oy 583 RDQIGLGWGSQDGTGR-----LVTKFEHNLINRGVQAGAE- 619
Db 248 ----IAGDEPITGTQRPQIYSGLARFQPGMPYDLDDLDFQOAL-EQNGHYSGASVQ 302
Oy 620 ----IASEDKKGVKLYATKPLSHPLNDOLRAT--LG-----YOQEVFGHSTNG 661
Db 303 ADFRLQGDVRPVKVSYTEVRKRLKETGIRLDSRYGLGKRIAYNYLNFKNIGSYVMD 362
Oy 662 FDLSTRTLEHISRSITIQNGM--NRTYSLRYRLDKLTKQAPPEWMDLPDVFNGKPS 718
Db 363 MDKETTLAGISQPRNRYRNWYTSNYSNRTQNLKRAFSGGV-----YVRDRAG 416
Oy 719 QEALLA-----GVAV-----HKT-----VADNLVNPARKY-----RQRY 747
Db 417 IDARLGAFLAEGRIKIPGSAVDLGNHATMLTSMKROLNNVLPBNGHYLQCKIGTTL 476
Oy 748 SLEVSSGLVSDANMALIARAGISGVYSGDNAYGSNRAHQMTGCI-----QAGYIWSDFN 803
Db 477 GFELSSYALIRTS-----ARAG-----YFFPENKKLGFILIRGAGYTVARDNA 521
Oy 804 HVPYRLRFAGDQSIKGYADSLSPISDKGYLTGQVLAAGTAEYNEFMKDLRLAVFG 863
Db 522 DVPSGLMFRSGGASVYRGYELDISLAGPNSVLPERALVGSLEYOLPPTRTLSGAVH 581
Oy 864 DIGNAVDCGFTNTKIGAGYGVWMAAPVQGVYDVAAGVKEEKNPKLHFFIGTPE 919
Db 582 DMGDAANFRKMKLKHGSGSLGVWMSPLAFSPFDIAYGSDK--KIRWHISLGTFR 635

```

## RESULT 28

09PEI2

PRELIMINARY: PRT: 784 AA.

```

AC 09PEI2
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Outer membrane antigen.
GN XF1046.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX Xylella.
RN NCBI_TaxID=23711;
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Artuda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Bata G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
RA Coultano N.B., Colombo C., Costa F.F., Costa M.C.R., Costa C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dority H.,
RA Facinani A.P., Ferreira A.J.S., Ferreira H.M.F., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kiteajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madelira A.M.B.N., Madelira H.M.F., Martino C.L.,
RA Marquês C.F.M., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.V., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,

```



RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
RA Zago M.A., Zatz M., Melandis J., Setubal J.C.,  
RT "The genome sequence of the plant pathogen *Xylella fastidiosa*."  
RL Nature 406:151-159(2000).  
DR EMBL: AE003941; AAF83856.1; -  
DR Interpro: IPR000184; Bac\_surfAg\_D15.  
DR Pfam: PF01103; Bac-surface.Ag; 1.  
KW Complete proteome.  
SQ SEQUENCE 784 AA; 87991 MW; 719331DD881E8192 CRC64;

Query Match 4.5%; Score 214; DB 16; Length 784;  
Best Local Similarity 20.5%; Pred. No. 0.00038;  
Matches 177; Conservative 100; Mismatches 300; Indels 266; Gaps 45;

QY 228 NSGIPRLQRTALVAARAVGYDIDLSIIRNSIEVDYIINDLEPEVYIDYRAVEVREGA 287  
DB 17 NPSLPVLTQAA-----ESFVANDI-RVDG-LQRIASGVITYLLPYN-RQDIV 60  
QY 288 DDKAFITTVADENVPLIGDVFHHG-----KYEYK 315  
DB 61 DD---AKVADAIKALYRTGFEEFNVRIDROGNIIVYKKEPRAINKLITGNKDIKSEEL 117  
QY 316 KNLIEAASHEGVGDRMLDRSDVLLPONTADVSLIYDTGTQYRDEVEVFTIDPRTNQ 375  
DB 118 KGLSEIGLSEGGTFDRSLDRV-----TQELKROYNNRGKYNV-----Q 156  
QY 376 LTTDPDLKPYKRELBOLITLVNNGEAYNLQAVRALSNDLATRYFNNVNTIEVPEREQ- 434  
DB 157 MTT--TTTPLDRNRVDVTLAIKKGAKAIRHI-----NLGTGKFNKKDQMSMESKEHN 209  
QY 435 -----IQNDQVS-----FEQSSSRTEPAQVD-----ESTLEP---VIEFTVELTDG 472  
DB 210 WASWYRRDDQYSKEKLSGDEKLSNWTYDRGYVDFNIDSTQVSIPEKHMMFTAGVTEG 269  
QY 473 ILMDISPIEFESANLI-OKL-NLVAKARHLDMEDDRVLAINHDDGVARSLTGRSDA 530  
DB 270 DQYKISSIVGTNTVLPQETKTEKVLTPKGTDI---SRFL-LEYSASAIINLNSIGYA 324  
QY 531 VSAVARAILPDESENEVIDLPERTALANRKTTPADVYQSKKVPV--VYFVASDPRDQIG 588  
DB 325 FSKVN---PIPTANRA-----DRTVAVN---LHVIPGRVTVYRQILFKNTTSD---- 368  
QY 589 LQMGSTGTRLTYKFEHN-----LINRDGYQAGAEIRL-----SEDKKGVKLY 631  
DB 369 -----EVLRRERKOFENSWTQQAIDRSKIRLQRLGYFEAVDESTPVPQSNQVDI-VY 422  
QY 632 ATKPLSHPLNDQRLATLGGQEVFHHST-----NGFDLSTRLEHHSITQ----- 679  
DB 423 YVKEIT--SGSFQVGLGY-SKTYGVTTVSQLSQNNFLGSGNRSVDSRSRYQDRYSFS 478  
QY 680 -----GNRNTY-----SLRYRLDKLKTQ--APPETMODLPEVFNCK 716  
DB 479 YNPFPTDNGVSLGYMLAQOKLDYDFPMAQVYSKRMSCGTTIGCITPTEHD-IVSWVIGA 537  
QY 717 PEOBALLAGVAHKTVAD-----NLVPMRGYRQYSLEY 751  
DB 538 DSNQITTFPFGSPKALIIDYDAVGQRTFPAWTELMQARDTRNDYDPMPLGMQIRIGAEV 597  
QY 752 ---GSS-----GLYSDANMAIARAGISGVSPEDNMGSRRAQMTG 790  
DB 598 TLPGSTIKYKKINYOISKWPIIPALVLTMLREV-----GYDDD-YGKSHRTILPD 647  
QY 791 G1OAGYIWSDNFNHVPYRLRFAGDQSGIRGYAHDLSLPSD-----KGYLNGGVQLAV 844  
DB 648 GYVA-----TASGLPFENNFAAGTNSVYRGFRDNTLGRSEVTALYNQOQPLGGSFKEY 701  
QY 845 GAEVYEEFMKD---LRLAVFGDIGNAYD--KGF-TNDTKIGAGVGVRRASPVGOVRVDV 898  
DB 702 GSTEMFFPLFDSPSARISAFIDFGVNFVGNVFNKANELBRASGVALLMRAPIGPISISY 761  
QY 899 ATGVKEEGNP--IKLHFFGTPE 919  
DB 762 APPIKKNENDEIERLOFTFGQF 784

RESULT 29  
Q8ZH58 PRELIMINARY; PRT; 795 AA.  
AC Q8ZH58;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Putative surface antigen.  
GN YP01052.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Yersinia.  
OX NCBI\_TaxId=632;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CO-92 / BIOVAR ORIENTALIS;  
RX MEDLINE-21470413; PubMed-11586360;  
RA Parkhill J., Wren B.W., Thomson N.R., Tlhalil R.W., Holden M.T.G.,  
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Feltwell T., Hamlin N., Holtroyd S., Jagers K., Kariyasev A.V.,  
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.,  
RT "Genome sequence of Yersinia pestis, the causative agent of plague."  
RL Nature 413:523-527(2001).  
DR EMBL: AJ414146; CAC89894.1; -  
DR Interpro: IPR000184; Bac\_surfAg\_D15.  
DR Pfam: PF01103; Bac-surface.Ag; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 795 AA; 87838 MW; 8C4581328D38F821 CRC64;

Query Match 4.5%; Score 211.5; DB 16; Length 795;  
Best Local Similarity 19.1%; Pred. No. 0.00055;  
Matches 172; Conservative 139; Mismatches 306; Indels 283; Gaps 42;

QY 148 NSEVWPPTLEPKPG-LIKRLVA-----RLFNDGVNKPRLKAKFYQSQSGERSAI 199  
DB 45 NMFVVGDIYSDDDIGKTIALPATGNEFDYRVLRDQNTLIVQK-----EKPTI 94  
QY 200 GSSHOKTEPYANIKAALEDITQESANDLNGSIPRLQRTALVAARAVGYDIDLSIIRN- 257  
DB 95 ASI-----TESGKAKAYKEDMLKQ--NLKASGYRVGEA-----LDPTTISNIE 134  
QY 258 -----STGEVDYIINDGEPV---YIDYRAVEVREGA-----DDKAFIT-- 294  
DB 135 KGLEDFEYYSYKASAKAVVTPLRNRRVDLKLVTFTGVSAKIQOINIVGNHSFTDELI 194  
QY 295 ---VADEVPL--LIGDVFHHGKYEYKKNLLENASAHEGVGDRMLDRSDVLLPONTAD 348  
DB 195 SRQLRDEVNMMWVVGQ---KKYOKKLAGDLETLSFYLDKGIAFRFNIDSQVSLTPD 250  
QY 349 VSLIYDT-----GTQYRDEVEVFTIDPKNLTTPDKLPVARELLEQLTVNMEAYN 403  
DB 251 KKGIVTYINTTEGPQPKNSVIY-----SGNLAGHQSSEKELTKIEPPELEN 297  
QY 404 LQAVRALNSDL--IATRYFNNVNTIEVPEREQIQNDQVSFEQSSSRTEPAQVDESTLE 461  
DB 298 GSKVLTMEDDIKKMLRGYGAUVRVVTQPE---IND-----DDKTVK 336  
QY 462 PVLETVELDGLIMDISPIEFESANLIQOKLNLVYAKARHLVD--MPDDVYLAINDDGV 519  
DB 337 ---LHINVDAGNFFYVRHILFESGNDTSKD--SVLRERMRQMEGAMIGNDQVE--GKERL 389  
QY 520 NRSILIRISDAVASAVARAILPDESENEVIDLPERTALANRKTTPADVYQSKKVPV- 575  
DB 390 NR--LGYF-----ETVDY-----EQGRVPGADIVD 413  
QY 576 FVASDKPRDG---QIGLQMGSDTGTRLVTKFEHNLINRDGYQAGAEIRLSEDKKGVKLYA 632  
DB 414 VYKYKERNTGSLNFGIGYGTESG---VSFGVGOODMMLGTGNTVVGINGTKNDYQTYA 469

```

OY 633 TKPLSHPLNDOLRATLTGVOEVEFGHSTNGFDLSTRTLEHEISRSIIIONGWM-NRTYSRLY 691
DB 470 EFTLMDDY-----FTYDGVSLGGRIFRYNDKADNADISGTTNSYGC--- 510
OY 692 RLDKLTQAPPEYTMODLPVDFVNGKPSOE--ALLAGVA-VHKTAVADNL-----V 737
DB 511 -----ADGTLGFPINENNSLHVGVGYNHNDSDMLPQVAMMYLESV 552
OY 738 NPMGYROR-----YSLEVG-----SSGLVSDANMAIARAGISGVY---SF 775
DB 553 GERGVYGRGEGFTTDDFTLNLGWTYNNLDKGFPTSGVSKSVNKITVPDGNDFEYVTF 612
OY 776 GDNAY---GSNRHOMTGGIOAGYIWSDNFNHVPYRLRFPAGDQOSTRGVAHDSLP--- 829
DB 613 DTSAYOPLNDRSMVLLGRRLGIGDGSKEPMFENFYFAGSGSTYRGFRSNNGFKAA 672
OY 830 -----ISDKGLTGGVAVGTAE-----YNEFKMDLRLAVFGDIGNAYDKGFT 874
DB 673 YVANGATVNSTDAVGNAMAAVASELILPTPFISKEYNSVTSIFIDSGTWTDTNME 732
OY 875 NDKR-----IGAGYGRMASPYGVAVDVATGYKE-EGNPT-KLHFFIG 916
DB 733 NTAKTRAGIPDYGKASNIRVSAGVALQWMSPLGLPVSFAKPYKDYEGDKSEQFOFNIG 792

RESULT 30
OBYE13 PRELIMINARY: PRT: 623 AA.
AC 08YE13:
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DE Outer membrane protein.
GN BME11895.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID:29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA Delvecchio V.G., Kapralov V., Redkar R.J., Patra G., Muij C., Los T.,
RA Ivanova N., Anderson I., Bhatlacharya A., Lykdis A., Renik G.,
RA Jablonaki L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Teleson J.-J.,
RA Haselkorn R., Kyripides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL: AE009623; AAL53076.1; -.
KW Complete proteome.
SQ SEQUENCE 623 AA: 67227 MW: 07C43D0F6ADBF8FE CRC64;

Query Match 4.4%; Score 209; DB 16; Length 623;
Best Local Similarity 21.0%; Pred. No. 0.00051;
Matches 146; Conservative 79; Mismatches 265; Indels 208; Gaps 28;

OY 310 GKYERKKULIENASA-----EHG-----YFDGRW-----LD 335
DB 42 GKADLKSIVIGASGLVSDADKPAAGSAGLLAKARGDYRLILSALYGCGRGGTISIKVD 101
OY 336 -RSVDVILPD-----NTAVSVLIDYGTQYRFDEVYFTIDKTNOLTTDPDKLPVKR- 388
DB 102 GREANDIPDPEIPNNNAVATVDPGQPLFSRAISNIAPPGRNR--DKVOTPEBAG 158
OY 389 -----LLEQLTVMNGEAYNLAQAVRAISNDLIATRYFMVNTETVLPFEREQ 436
DB 159 FAPGGEAKSCTILAEKRLAVAMNOEGAKARVGEDVAVHADRRVSADIALDPGRKAY 218
OY 437 NDQVSEFOSSSSRPEAPQVDEST-LEPIETVELTLDGILMDISPIEFASNIQDKLNLV 495

```

```

DB 219 YGPVSV--VGTRAMDPOFVAMMTGLKPEQE----- 246
OY 496 AAKARHLYMDPDDRVLAINHDDGVNRSLIGRISDAVSAVARAILPDBSENEVIDLPERTA 555
DB 247 -----YD-PDDIENM-----KKRLGRME-----VFRAAMFEEADK----- 275
OY 556 LANKTADVYQSKKVPVLYFVASDKPRDQIGLGMSDGTGRLVTFEH-NLINRDQYQ 614
DB 276 -----LEPDGSLPITLNVQERKPRRGFGAEYSTIDGFGVTSYWMHNNLLGR----- 322
OY 615 AGAELRLSEDKGY-----KLVA-----TKPLSHPLNDOLRATLTGVOEVEFGH 657
DB 323 -GERLRPDADAVSGISGSDNSFPKNTTYLLGASFAKAPGYTTPDPTDVAITLDAREV--- 378
OY 658 STNGFDLSTRTLEHEISRSIIIONGNMRTYSLRYRLDKLTQAPPEYTMODLPVDFVNGK 717
DB 379 -----LDAYET-----SINAKGTQIRS-----DELSGALYANNSQGHFVDVFGK- 421
OY 718 SQEALLAGVAVHKTAVADNLVPMRGYRQRYSLYEGSSGLVSDANMAIARAGISGVSE-- 775
DB 422 -RQFTTAG-----LEGNLT-----YDSRNKKPDSSGFFYLVGN-----IQPFYEFH 462
OY 776 GD-----NAYGSNRHOMTGGIOAGYIWSDNFNHVPYRLRFPAGDQOSTRGYA 823
DB 463 GNFAFRPTAGRTYHGFQIDRVYLAGRLKVGSIIGSIADLPSSQLFLAGGGGVNRYG 522
OY 824 HDLSPLISDKGLTGGVAVGTAEYNEFKMDLRLAVFGDIGNAYDKG---FTNPTKIG 880
DB 523 YRNIGVSAGNGNIIIGRSLVEANGEVTRITDLSIGAVAFADAVYGEKSPDSEDMRVG 582
OY 881 AGGVAVMASVGVGRVDVATGYKEEGNPKLHFFIG 916
DB 583 VGGGLRYLTSLGPIRLDVAVPLNRSGDPNRYGYYG 618

RESULT 31
OBYE13 PRELIMINARY: PRT: 916 AA.
AC 025369;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Protective surface antigen D15.
GN HP0655.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervatave A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RT Nature 388:539-547(1997).
DR EMBL: AE000579; AAD07720.1; -.
DR TIGR: HP0655;
DR InterPro: IPR000184; Bac_surfaq_D15.
DR Pfam: PF01103; Bac_surface_Ag_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 916 AA: 102671 MW: 186D8E3D1C09408A CRC64;

Query Match 4.4%; Score 206.5; DB 16; Length 916;
Best Local Similarity 19.3%; Pred. No. 0.0014;

```

Matches 202; Conservative 143; Mismatches 318; Indels 383; Gaps 51;

QY	24	MSQALAOONNPANIINIHNPAIDTALINAKAANPPVYLITPEOIOAR-LNAAGLNAKPOSQ	82
Db	31	LTPSKAEASOBQSK--NEAPKKEVORNEKQETPOSNQTPKEMKKSISYGLSYMSDML	87
QY	83	ALDVNFDDQSPISRGESOPPLGIDMSVIEETPTLSLEELFAOESTBMGINPNDIPEY	144
Db	88	ANEIVKI-----RQVDIVDSKKIDTIANVL-----ALFNO-----	115
QY	143	QGBQNPSEVVPPTLEPEKPGILKRLYARLENDGVNKPRLKAFYOSSQSG--ETSAG	200
Db	116	-----GFKQDVYA--TEEGGI-----LEFHEDEKARIAGVEINGYG	149
QY	201	SSHQKTEPYANIKAALEDITQESADNLGSIPLRQOTAL--VAARVGYDDIDLSIINSI	255
Db	150	TEKEREKGLKSQMGIKKGQJFDEQAKLE-----NAKTLKLTALGEGY	191
QY	260	GEYDVIIHDLGEPYIIDYRAVEVEBEGADKAFYTVAD---EVPLLIGDVFHNHKEYTKR	318
Db	192	GSV-----VEPTEKVSSEGALLIVPVRNGDSIYIKOSIYESAALKR	234
QY	317	NIENNASAEH-----GYFDR-----WLDNSVD--VILPD	344
Db	235	RMEESIANKKODKPMGMMGLNDGKLRIDOLEYDSMRIQDYVMRGYLDIAHISSPFLKTD	294
QY	345	-NTADVSLIYDT--GYORFEDVEVFYITDPKTNQTTDPDKLPKPRELLPOLLVNNGEA	401
Db	295	FSTHQAHLKHYKKEGIOYRISDIL-----IENDPNVPLK--TLEKALKVKKDY	342
QY	402	YVLQAVRALSNLQ--IATRFNNMVEIYEPREKQIONDOVSFEQSSSRTEPAQVDES	458
Db	343	FNIEHRLADAQALKTLEIDKGYAFA--VVPKPDLDK-----DEK	378
QY	459	T-LEPIETVELTQDILMD--ISPIEFSASNNL-----QDKLNVAAKARHLYDM	506
Db	379	NOLVAVIYRIEGBDVYIINDVYIISNOQTSRITIRRELLGPKKQKWLTYLR-----	430
QY	507	DDRVLIANHDDGVNRSILGRISDAVSAAVARAILPDESENEYIDLPERTALANKRTPADY	566
Db	431	-----NSENSLRR--LGPFSSKVIIEEKRY-----NSSLMD-----	459
QY	567	QSKKVPVLYVAFAASPPROGOI--GLGMSDQITGLYAKFEHNLINRDGYOAGELRLSED	624
Db	460	-----LVSEBGRGTQLOLFGIGSSYSGML-----NGSVSEHMLFQTGOSMSILYAN	506
QY	625	--KKGKLYATKP-----LSHPILNDQLRATLIGYQOEFHGSHTGEPDLSRTRL--	669
Db	507	IATGGGRSYPCGMPKAGRMFAGNLSLNP-----RIF-----DSWISSSTINLY	549
QY	670	-EHEISNATIONG-----WNRTY--SLARYLDKLTQAOPPETWODLPYDFPNC--	715
Db	550	ADYRISYIOIOGGGFGVYNGRMGLNRTHTSHLIGNLVTKILGFSSPLYNRYSSVNEVY	609
QY	716	KPSQEFALLGAAVHKHTVA-----DMLVNP-----MRGROR-----YSELY	751
Db	610	SPROGTSASYIINRLSGKTPLOPECSBGALITTSPEIRGLIMDRDHYHTPITSSPELTV	669
QY	752	G-----SSGLVSDANMAIARAGISGVSPGD--NAYGSRARHOMTGG-----	791
Db	670	SYDWTDDYFFPRNGVIFS-----SYATMSGPSSGTLNSMNGGLGAVNRMTKYGKTAH	724
QY	792	-----IQAGYIMSNF--NHVPYRLRFPAGDOSIRKYAHDLSLSPISDKGY	835
Db	725	HLQGYLLIDLIAKPKTGGYIFRINTDDYJPLNSTEYMGVTVRBGFRNGSVTPKDEFG	784
QY	836	LTTGGOVLAVGAENYEFMK--DLRLAVPFDIG-----N	867
Db	785	WLGGOGIYFTASTELSYVLAKRAAKRLAMFDFGFLFKPTPTRSSEFYNAVPYTTANFKDYG	844
QY	868	AYDKGFINDT--KIAGYGVNRMASFVG	892
Db	845	VIGAFERATWRASTGLOIMWISMG	870

```

RESULT 32
Q9ZE03
ID Q9ZE03 PRELIMINARY; PRT; 768 AA.
AC Q9ZE03;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Outer membrane protein OMP1 (OMP1).
GN RPI60.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichteritz-Ponten T., Almark U.C.M., Podowski R.M., Neeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
DR EMBL; AJ235270; CAA14627.1; -.
DR InterPro; IPR000184; Bac_surfGp_D15.
DR Pfam; PF01103; Bac_surface_Ag1; I.
KW Complete proteome.
SO SEQUENCE 768 AA; 87196 MW; B1AB12D783DF9DCB CRC64;

```

Query Match	4.3%	Score 205.5;	DB 16;	Length 768;
Best Local Similarity	20.1%;	Pred. No. 0.0012;		
Matches 171;	Conservative 111;	Mismatches 297;	Indels 271;	Gaps 40;

QY	165	IKRLVAR-----LJNDG-----VKNVPLKAKFFUOSSOGSETSAI-----GSS	202
Db	65	IKRLVATLEFRNINMYITNDGMLIVNTEPEPISVSFSGNSKIKTNILAKEIYTMSSGS	124
QY	203	HOKTEPYANIKAALEDITQESAMDUNGSIPLRQJALVAARAVGYDDIDLSIRNSIGEV	262
Db	125	LSQAKIEIDVKKILE-IYKRSGRFETKATPKK-----SLENNKRV---	163
QY	263	DVLIHIDGEPVYIDYRAVEVRGEG--ADDKAFYTADEVP-----LLIGDYVHHGKEYTK	315
Db	164	-KVJFDIAGEGRTVAKSLIYFSGNEHYISDSLSKLYLTRESMRFRLESMDITYDPPRVEID	222
QY	316	KNLIEMASAEHGYEDGRWLDRSDVILIDPMTADVSLIY--DTGYQYRDEVEVFFYIDPKT	373
Db	223	KELLREFYOSVGFADERYI--SASVALNDTKREFYITYSIEGEGEYRRGNV---IID---	274
QY	374	NQLITDDPKLPVKKELEQLLTVNMGEAYNIOANVALSNDLITATRYFN---MYNTEIV	428
Db	275	NKLTN-----INIKOLKIVINIKOGKIFEMKKTVDIDIAEKI--GEFYANGAPYAVN--V	323
QY	429	PPEREOJONDVSEFOSSSSRTPEAQVDESLPEYIEVELTDLGIMDISPIEFASMLI	488
Db	324	YPD-----IKRNDNHTADIKPEIIEKADYV	348
QY	489	QDKLNLVAAKARHLIYDMPDDRVLAINHDGQVNRSLIGRISDAVSAVARALLPDESENEVI	548
Db	349	INKNINII--NNLKREDNHYIRAFKIEEDVYNNRSYIEK-----GEKNLNL	392
QY	549	DLPERTALANKETPA-DVYOSKAVPLVYVFAVADKPRDQIGLWGSPTGTRLVYTF--E	604
Db	393	DYFEKVASISLAOTKAKDKYD-----VNVVEDKSSISIGFDLGYNAGGLFGRFSLE	445
QY	605	HNLIINRCYOAGALRLISEDKKGYKLATKP--LSHPLNDOLRATLGY-----Q	651
Db	446	RNLVIG--TGKLLINAGVOYKSNSTSYGCGITDPHPHFDRLDLSVNAFRNNTYTGASVALLTTD	504
QY	652	QEVFGHSTN-----GFLSLRTLEHE-----ISREITIONGGMNRRYSLRYRLDKLAKTOAPP	702

```
Db 505 OSYKLSHGVKISLGYDNK-EDLSHEIDYLIKRDLISAPSPSNSIFLNEOMKLTITSA-- 561
QY 703 ETWODLPVDFVNGKPSOEBALLAGVAVHKTVADNLVNPGRYRQRYSLSEVSSGLVSDANM 762
Db 562 -----ICHTITTDQTDNKTIVPKNGTL----- 582
QY 763 AIAAGISGVYSF-----GDNAVGSNRAH-----OMTGGIAG-----YI 797
Db 583 -----VSGTGEFAGVGDGNKYIKHEIECKFYKSFINKKVTLLKLSAAGDMAAGCGKKMVR 636
QY 798 WSDNNHVPYRLREFAGGDOSIRGVAHDSLSF-----ISDKGLTGGQVLAVGTAENY--- 851
Db 637 ISDRFNL-----GDYSLRGFASGCVGPREKNTNEG--LGGERYYTESTELNPTP 684
QY 852 ---EFMKDLRLAVFGDIGNAY-----DKGFTNDTKIGA--GVGVWASPVGQVR 895
Db 685 VPREF--NFTGAVFLDGLGSVGVGLNKKQYKTPNGFYNDOSLRASVSGGFITWTRFAPIR 742
QY 896 VDVATGVKEE 905
Db 743 MDWGFPIKKK 752

RESULT 33
ID 092J67 PRELIMINARY; PRT; 768 AA.
AC 092J67;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Outer membrane protein ompl.
GN OMPL OR RC0202.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
NCBI_TaxID=781;
RN [1];
RC SEQUENCE FROM N.A.
RC STRAIN=MALISH 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fourmter P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT Mechanisms of evolution in Rickettsia conorii and R. prowazekii.;
RL Science 293:2093-2098(2001).
DR EMBL: AE008587; AL002740.1;
DR InterPro: IPR000184; Bac_surfag_D15.
DR Pfam: PF01103; Bac_surface_Ag; I.
KW Complete proteome.
SQ SEQUENCE 768 AA; 86881 MW; F0327FC56408E36A CRC64;

Query Match 4.3%; Score 202.5; DB 16; Length 768;
Best Local Similarity 20.1%; Pred. No. 0.0017;
Matches 172; Conservative 122; Mismatches 301; Indels 261; Gaps 43;

QY 164 LKRLYA-----RLFNDG-----VNVKVPRLKAKFYOSSOSGETSAI-----GS 201
Db 64 VIKRLYATSLFRINNHITNDGLIVNVTETPFISSVSVSGNSKITMTMLAKETITMSGE 123
QY 202 SHOKTTPYANIKRALDITIOESAMDINGSLPRKROTALVAARAAGYVYDIDLSIRNSIGE 261
Db 124 SLSQAIEIDVKKILIE-IYKRSGRFATVYTPKIE-----NLNNRV-- 163
QY 262 VDVYIIDL--GEFVYIDYRAVEVRGEGADDKAFTTVADEVP-----LLIGDVPHHKYER 314
Db 164 --KVIFDIAGEGRTGINTYFSGNENYSDSELKSLVLTESRMFRLESNDTYDPRMEY 221
QY 315 KKNLIEMASAEHGYEDGRWLRSVDVILPDNTAD--VSLIY--DTGTQYRFDEVEVFTIDP 371
Db 222 DKELLRDFYOSVGFADFVRISASAEI---NNTKEYFTLYTIEEGKYSFGNI---TIDN 275
QY 372 KKNQLTTDDKLPVKRELLEQLITVNMGEAVNLQAVRALNDLIATRYN---MVNTE 426
```

```
Db 276 KLTNIDITP-----LNKIVNIKQGVFNMMKVVDDIAEKI--GEFTANGYPAVN-- 322
QY 427 IYVPEREIQNDQVSFEQSSSSRTPEAOVDSTLEPVIEVELDGLIMDISPIEFSASN 486
Db 323 -VYPN-----IKNNNHT---ADIKFITEKADK 346
QY 487 LIOOKLNLVAAKARHLVMPDPRVLAINHDDGVNRSILGRISDAVSAVARILDESENE 546
Db 347 VYIKINII--NLKTEDPVIRREKIEGILINRYTEK-----GERNLR 390
QY 547 VIDLPERTALANRTPA--DVYQSKRVPVLYFVASDKPRDQIGLGWSDGTGRLVTKP-- 603
Db 391 NLDFEKAVALSLAPTKAKDKYD-----VAVDEKSTSSIGDGLGYNTAGGLFGRRSE 443
QY 604 -EHNLI-----NRDGYQAG-----AELRLSEDK-----KGVKLYATK 634
Db 444 LERNLVGTGKLLNAGVOYSKNTSYGGITPAHPHFLDRSLSGVNTFRNYTGRGASVLTNT 503
QY 635 PLSHPLND-QLRATLGYO-GEVFGSTNGPDLSTRILEHISRTIONGNNRTYSLRYR 692
Db 504 DQSTKLHSGVKTSLGTEIKEDLGE-----IDYLIKRDLISAPPPSSIFLKEQ 553
QY 693 LDKLKTQAPPETWODLPVDFVNGKPSOEBALLAGVAVHKTVADNLVNPGRYRQRYSLSEVG 752
Db 554 MGRFITS-----ICHTITTDQTDNKTIVPKNGTL----- 590
QY 753 SSGLYSDANNAIARAGISGVY--SFGDN-----AVGSNRAHOMTGGIOAGYI--WSDN 801
Db 591 GVG---GNKKYIKHEVDGKYKSFINKLTLKLSAGGN---ITG--LGKILIRISDR 640
QY 802 FNVHVPYRLRFAGGDOSIRGVAHDSLSF-----ISDKGLTGGQVLAVGTAENY-----E 852
Db 641 FNL-----GDYSLRGFANGCVGPREKNTNEG--LGGERYRKSTELNPTPIPEE 688
QY 853 FMKDLRLAVFGDIGNA-----YDKGFTNDTKIGA--GVGVWASPVGQVRDVA 899
Db 689 F--NLGTGVFMDLGSVGVGLNKKQYKTPNGFYNDTSLRASISGGFITWTRFAPIRDMG 746
QY 900 TGKVEEENPKIKHFTI 915
Db 747 FPIKKKKYDDTOHFL 762

RESULT 34
ID 09X3V8 PRELIMINARY; PRT; 738 AA.
AC 09X3V8;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 81.1 Kda protein.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
NCBI_TaxID=542;
RN [1];
RC SEQUENCE FROM N.A.
RC STRAIN=ZM4;
RA Lee H.J., Kang H.S.;
RT *Sequence analysis of 42C11 fosmid clone of Zymomonas mobilis ZM4.*;
RL Submitted (Aug-1998) to the EMBL/Genbank/DBD databases.
DR EMBL: AF088896; AAD21534.1;
DR InterPro: IPR000184; Bac_surfag_D15.
DR Pfam: PF01103; Bac_surface_Ag; I.
KW Hypothetical protein.
SQ SEQUENCE 738 AA; 81140 MW; F169545D09168114 CRC64;

Query Match 4.2%; Score 198; DB 2; Length 738;
Best Local Similarity 20.6%; Pred. No. 0.003;
Matches 176; Conservative 122; Mismatches 344; Indels 212; Gaps 45;

QY 127 ESTEMGINNDYI---PEYGEQPNSEVVPPTLEPEKPGLIKRLYALFNDGVNKKVPR 182
```

```
Db 36 DSGNKAASNPNAIDPFDPSPFDHLP---TLSPSTSEASKA-----DSSSHLP- 79
QY 183 LKAFYSSOGSETS-----AIGSSHQTEPYA-NIK-----AALEDI 219
Db 80 --LRTSESSOGSENSNGKGVKALDTEVSDPKNNLANSDBEKTAKQIEETSTYLDPA 137
QY 220 TQESAMDLNGSIPRLRQTALVAARVGY-YIIDLSIIRNSIGEVDVITIHGEPYIDYR 278
Db 138 LSEPLPLDSHPMMEHVNLAKESDDLRYKVYKGLSN-----IDT-----DQAFKYS 187
QY 279 AVEYRGEGADDPKATYVADEVPLLIGOVFHHGKETKNLLENASAEHGYEDGRRLDSV 338
Db 188 VLISNNKKAASLS-----VIG-----AKATNDELINRLIRSGGYDG---KAA 228
QY 339 DVILPDNTADVSLYD--TGQYRFDEVEVFPTIDPKTNQLTDPDKLPVKRELL-----E 391
Db 229 LSTPIEKGQYVNVRYDNAGSIYKLGQI-----NLTQSEEPRLRIARVALNLHNGD 279
QY 392 QLTVMNGEAYNLQAVRALSNDLIATRFENNVNTEIVFEREQIQNDQVSFEQSSSRTE 451
Db 280 PILATHEQAEH-----NILITLPHYGPFAKI--GDRITILDD---ETHGQVTL 325
QY 452 PAQV-DESTLEPIETVELTQGLIMDISPIEFASNLQDKLNVAANKARHLYD--MPDD 508
Db 326 PVNAGNVSGSIT--VSNNKHIVLDKAKHS-----HIAKREGQRYDSQWVD 372
QY 509 RVLAINHDDVNRSLIRISDPAVASAVARAILPDESENEVIDLPERTALANKTPADVYOS 568
Db 373 LRQAL-----AATSLFHSVSEPIATGR--KREDGSEYDL-----DVRQG 411
QY 569 KKPPLVYFVASDKPRDQOIGLWGSDTGTRLVTKFEHNLINRDGYQAGAEIRLSEDKGV 628
Db 412 R-----GKKHSIAYVGYGTGEG--FKAQGSWISRNYPPEGSLTFS---GI 453
QY 629 KLVATKPLSHPLND-----QLRATIGYQOEYFGHSHNGDLSRTLIEHSISIT 678
Db 454 LGTRQOOLSLFNRNNAGANDRYIQIGTAGREKY---DAVNGSFS---LGSSLSRQST 507
QY 679 QNGGWNH--TYSLRVRLDKLKTQAPPETWODLPYDFVNGKPSQE---ALLAGVAHVKT 732
Db 508 QL--WQKRMVYISIGAL--TOTNERSY-----DFSRSELTFTFYLLAALPGQLGYDR- 555
QY 733 ADNLVNPARKYR--QRYSLFVG--SSGLVSDANMAIARAGISGVYSGFDNAVGSNRAHQMT 789
Db 556 SNNLMPYTKGYRLNMRSPETSISGLRGYVRLMF--DASGYVPADVNV-----LA 605
QY 790 GGIQAGYIMSDNNHVPYRLRFAGGDSQSTRGAHDSLSPSDKGYLTGGOVLAVGTAEY 849
Db 606 GRRVVASIEGASVOELAPSRRTIYAGGGSVRGYQDLGPKDPYKDAVGGISEBELAFEV 665
QY 850 NFEFMKDLRLAVFGDIGNAYDKGFT--NDTKIAGVGVWRMASPYGVQVNVATGVKKE-- 905
Db 666 RYRF--GNFGLVPRIDAGQVEDAIPFPHNLRFTGIGLCARITYTAAGPRRIDLAPIARQPG 724
QY 906 GNPFKLHFITGTPF 919
Db 725 ESRISYIISIGQAF 738

RESULT 35
QY0YUR6 PRELIMINARY; PRT; 833 AA.
AC Q8YUR6;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein Atr2269.
GN ALR2269.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxId=103690;
RN [1]
RP SEQUENCE FROM N.A.
```

```
FX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Matanabe A., Iriuguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001)
DR EMBL: AP003588; BAB73968.1;
DR InterPro: IPR000184; Bac_surfTag_D15.
DR Pfam: PF01103; Bac_surface_Ag_2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 833 AA; 89633 MW; 7F52B1E52FEA1B59 CRC64;

Query Match 4.2%; Score 197.5; DB 16; Length 833;
Best Local Similarity 19.7%; Pred. No. 0.004;
Matches 134; Conservative 104; Mismatches 255; Indels 187; Gaps 30;

QY 371 PRTNQLTDPDKLPVKRELLBOLLVNMGEAYNLQAVRALSNDL---IATRFPMNVNTEI 427
Db 210 PNATPETTEP-----KLVSEVL-----VAPQSGQLPELETQVNVYIKTQ- 250
QY 428 VFP-----EREQIQNDQVSFEQSSSRTEPAQVDESTLEPIETVELTQGLIMDISPIEFS 483
Db 251 --PGRTTTRSQLQEDINAIETGTFSSNVQASPEDTPLGVRSFIVQRPVL---SKVEIQ 305
QY 484 ASNLQDKLNVAANKARHLYMPDDRVLAINH--DDGVNRSLIGRISDPAVASARAI-LPD 541
Db 306 ANPGTNVPSVLPQATADEIFRAQYKILNLRLDQEGI-KELTKRYQDQGVYLVANVGAPQ 364
QY 542 ESENEVIDLP-----ERTALANKTPADVYQSKVPL-----YVPASAKPRDQGI-- 587
Db 365 VSENGVYTLQVABGVENISVFRNKGQDVNEGQPIRGHTQDIITREVELPFGQVFN 424
QY 588 -----GLWGSQDTGTR 598
Db 425 RNTVQKDLQRYFGTGLFEDVNVSLDPGTDPYKVVNVVYVRRSSGSTAAGAGISSSGLF 484
QY 599 LVTKFEHNLINRDGYQAGAEIRLSEDKKGVKYATKP-----LSHPLNDQLRATL- 648
Db 485 GTVSYQOONLNGRMOKLGAEVQGERELFDLRFTDPMWIGDPYRTSYTANIFFRRSISL 544
QY 649 ---GYQGEV-----FGHSTNGPDL--TR-----TLEHESRSLIQNGGNRTYSLRYR 692
Db 545 IFDQKDDIDRTFDPGNDPNDTNGQDRPVYTRLGGSVTFTRPLSANPFRAEWTASAGIQYQ 604
QY 693 LDKLKTQAPP-----ETWODLPYDFVNGKPSQEALLAGVAVHKVADVADNLV 737
Db 605 --FVSTRDADAGNLKRDCAVFDQDNKRTSEIYPLSF--SGTGEDULLVLQGAQRDLRNPL 661
QY 738 NPMRGYRQRYSLFVSGSSGLVSDANMAIARAGISGVY-----FGDNAVGSNRAHQMT 788
Db 662 QPTSGSFLRFG--VDQSVPVGSGNIFLTR--FRGSYSQYLPVKRTGP-----SKGPETI 711
QY 789 TGGIQAGYIMSDNNHVPYRLRFAGGDSQSTRGAHDSLSPSDKGYLTGGOVLAVGTAE 848
Db 712 AFNIQGGTVLGD---LPRYEAFTLGGSNSVGRY-----EEGALGSGRSFVQASVE 758
QY 849 YNFEFMKDLRLAVFGDIGNAYDKGFTNDTK-----IGAGVGVWRMASPYGVQVNVATGV 899
Db 759 YRFPVSVVSGALFDFVGS--DLGTSTRTAEVLNKSQSGYGLGVAVVOSPGLIRIDY- 815
QY 900 TVKKEGNPIKLHFITGTPF 919
Db 816 -GINDGDS-RINFGIGERF 833

RESULT 36
QY0ZQ48 PRELIMINARY; PRT; 776 AA.
AC Q9ZQ48;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
```

DT 01-DEC-2001 (TReMBLrel. 19, last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)  
DE Putative outer membrane transmembrane protein.  
GN OMP OR R01502 OR SMC02094.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396507; PubMed=11481430;  
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
RA Bolstead P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,  
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gallibert F.;  
RT "Analysis of the chromosome sequence of the legume symbiont  
Sinorhizobium meliloti strain 1021.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
DR EMBL: AL591787; CAC46081.1; -  
DR InterPro: IPR000184; Bac\_surfAg\_D15.  
DR Pfam: PF01103; Bac\_surface\_Ag; 1.  
KW Complete proteome.  
SQ SEQUENCE 776 AA; 84520 MW; CE261E6372A94726 CRC64;  
  
Query Match 4.2%; Score 197; DB 16; Length 776;  
Best Local Similarity 21.2%; Pred. No. 0.0038;  
Matches 165; Conservative 114; Mismatches 315; Indels 186; Gaps 41;  
  
OY 223 SAMDNGSIPRLROTALVAARAGVY-DIDLSIRNSIGEVDYIHD--LGEFVYIDYR 278  
DB 70 SMDIDASVRL-----YATGFSVDSTIV--SGSLVSVSENLQNVFNGNR 118  
OY 279 AVEYRGAGADKAFITVADVPILIGVFNHKGKETKKNLEMASAEHGTGDMDRSV 338  
DB 119 KIK-----DCKLQGVVTRQ-PL--GPYSVAIVETDQAIQDAVVAIGRSD--VTVTT 165  
OY 339 DVI-LPDMTADVSLIYDTGTQYRFDEVVFF-----TDPK-----TQOLT 377  
DB 166 QVPIAGRVNLAFFVINEGERTKITQINFVGENEYSDGRLOSVATATESGIFSLTKDV 225  
OY 378 TDPDKLPYKRELLBQLITVMKEAYNLQAVRALSNDLIATRYFMVTEIVFPEREION 437  
DB 226 YNPDKLRADDELLRQFY-YNRGYA-----DQVVSSEALNATNEYT 267  
OY 438 DOVSFEQSSSTREPQOVDSSTLEPIVETVELDGIIMDSPIEFSSNLIOCKLNIVAA 497  
DB 268 VTTITEGPRYDFGPVNI-ESTVEG-IDAEELR-GLVQSRGTVYKAKD-IQSTMSISK 323  
OY 498 K-ARHLVDMFDDRVLA1NHDDGVNRSILGRISDAVSAVARALP----- 540  
DB 324 RVASEGY--PFAVTPRGNDLANHTI--AVDVLVDOGEAIVYERIEICNTTTRDYVIR 379  
OY 541 ---DESENEVIDLPERTALANRTPADVYQSK-----KVLVYFVASDKPRDG 585  
DB 380 REEDVEGDAFN-QEWMARAKRLLEALGYFSSVNISTQPSAADRVYIVVDQDQSGSF 438  
OY 586 QIGLGMSPTGRLYTKF---EHNLLNRQYOGAGELRS---EKKGYKLYATRP--L 636  
DB 439 GIGAGISACDGGGFLVEASIEEKNFLGRQY---IRLAAGKEDSQYVNSFTEPYFL 493  
OY 637 SHPLN---DQLRATLGYOQEVFCHSTNGFDLSTRTLEHESRSIIQNGGNRTYSLRYRL 693  
DB 494 GYRLAAGFLFNENDFDDDNYSYNDGFSL-----KVTAPITE---NSTTLRYNY 542  
OY 694 DKLKTDQAPETMODLPVD-FVNGKPSQEALLAGVAVHKTVDANLVNPMRGYRQRYSLYEV 752  
DB 543 TELEVFEGDDELSS-PYDRVIDQSPWTRSSISQSYTYNLT-DDAQQLPHGILASVYQDEFA 600  
OY 753 SSGLVSDAMMALARAGISGVYSGDNAYNSNRAHQHTGCI--QAGYIMSDN-----FNHV 805  
DB 601 GLGGSDFYKLTGKA--KWYTYLHDEA-----DIIGSLSGSAGHLFETGSLLEVFDQ- 650

OY 806 PYRLRFAGGDSITRGVAHDSLSPISDKCYLGGQVAVGTAENYERF-----KDLRLA 860  
DB 651 -----FQLNSNDIRKGFERNNGDNLGGTTYETTASAEATFPLPGLRPSGSGFA 704  
OY 861 VFGDIGNAYDGFNTDFIKAGVGVY-----WASPGGVRYDVATGVKEE 905  
DB 705 LFVADGTLYG----NDVEIGGESYRGDNASLRASVGYSLWASFPGLRDYAVPAVKE 760  
  
RESULT 37  
09ZLH9  
ID 09ZLH9 PRELIMINARY; PRT; 906 AA.  
AC 09ZLH9:  
DT 01-MAY-1999 (TReMBLrel. 10, Created)  
DT 01-MAY-1999 (TReMBLrel. 10, last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, last annotation update)  
DE Putative outer membrane protein.  
GN JHP0600.  
OS Helicobacter pylori J99 (campylobacter pylori J99).  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
OX NCBI\_TaxID=85963;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99120557; PubMed=9923682;  
RA Alm R.A., Ling L.-S.L., Noir D.T., King B.L., Brown E.D., Doig P.C.,  
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,  
RA Tummino P.J., Carnuso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
RT Trust T.J.;  
RT "Genomic sequence comparison of two unrelated isolates of the human  
gastric pathogen Helicobacter pylori.";  
RL Nature 387:176-180(1999).  
DR EMBL: AE001492; AAD06181.1; -  
DR InterPro: IPR000184; Bac\_surfAg\_D15.  
DR Pfam: PF01103; Bac\_surface\_Ag; 1.  
KW Complete proteome.  
SQ SEQUENCE 906 AA; 101438 MW; DE81214E8452CD29 CRC64;  
  
Query Match 4.1%; Score 194.5; DB 16; Length 906;  
Best Local Similarity 19.9%; Pred. No. 0.0069;  
Matches 192; Conservative 131; Mismatches 311; Indels 333; Gaps 49;  
  
OY 116 TPLSLELFAQEST-----EMGINPNDYIP-ETV-----GEOPNSEVYV 153  
DB 37 TPKEAQKNEAONETSOSNQTPKEMKYKSISYVGLSYSDMLANEIAKIRKQDWDVSKID 96  
OY 154 PPTLEPEKPGILIRLYARLPNDGVNKPRLKAKFYSSQSG--ETSAGSSHQKTEPYAN 211  
DB 97 TAYVALFNGCYFQDYTA-TFENGI-----LEPHDEKARIAGVEIKGYTEKEDGAKSQ 150  
OY 212 IKAALBDITQESAMDNGSIPRLROTAL-VAARAVGYIDLSIRNSIGEVDYIHDLG 270  
DB 151 MGIKAGGTDEQKLE-----HAKTALKTALLEGQGY-----GSV----- 184  
OY 271 EPIYIDYRAVEYVREGADKAFITVAD--EVPLLIGDVFNHGKYETKKNLEMASAEH- 326  
DB 185 -----VEVTERKVSSEGLLIVFDVNRGDSIYIKQSYIEGSDKLRKRVIRISLANKQ 225  
OY 327 -----GYFDR-----WLDNSVD--VLLP-NTPADVSLYID 354  
DB 236 RDEMGMMGLNDKLLDQLEYDSLRIDQVYMRGYLDLHISSPFKTGFSTHDALHYK 295  
OY 355 T--GTQYRFDEVVFFITDEKTNQLTDPDKLPYKRELLBQLITVNNGEAYNLQAVRALS 412  
DB 296 VKEGIDYIRISDLI-----IETDNPVPLK--TLKALKVKKKDVFNIEHLPAADQ 343  
OY 413 DL---IATRYFNKVNTEIVFPEREQIONDVSFEQSSSRTPEAQVDEST-LEPVIETVE 468  
DB 344 ILKTEIADKGVAFA--VVKPDLK-----DEKNGLYKVIYRIE 379  
OY 469 LVDGIIMDSPIEFSSNLIOCKLNIVAAKARHLMDPDR---VLA1NHDDGVNRSILG 525

```

Db      380 VGD--WHTINDVITISGORTSDRI-----IRRELLGPKKYLTKLRNSNSLR--LG 430
Oy      526 RISDAVSAVARAILPDESENEVIDLPETALNAKRTPADYQSKKVLVYFVASDKPRDG 585
Db      431 FFSKVIEEKRV-----NSLMDL-----LVSVEERTG 459
Oy      586 QT--GLGKMSDGTGLVTFEHLINRDGYQAGAEHLRSD--KKGVLYATRP----- 635
Db      460 QLOFGIGYGGYGLML-----NGSVSERNLEFGTQSMSLYANIATGGGRSYPGMRKAGRM 515
Oy      636 -----LSHPLNQLRATLGQOEVEFGHSTNGDLSRTL--EHEISRSIIQNG----- 682
Db      516 FAGNLSLTNP-----RIF-----DSWYSTINLYADYRISYQYIQGGGGGVN 558
Oy      683 -----NNRTY--SLYRLDKLKTQAPPETMODLPVDFVN--GKPSQEALLAGVAVHK----- 730
Db      559 VGMGLGNRRHVSLGYNLNTKLLGFSSPLNRYYSVNEVASFRQCSPTASVLIINRLSG 618
Oy      731 ---TVADNLVNP-----MGYRQR-----YSLEVG-----SSGLVSDA 760
Db      619 RPLVPESSGSPGAIITTSPEIKGIMDRDYHRTITSFTLDVYDNDYFPRNGVIFS- 677
Oy      761 NMAIAGISGVYSPD-----NAYGSNRAHQMTG-----IQAG 795
Db      678 ---STATMSGLPSSGTLNSMNLGGVNRNTKYVKFAAYHLOKYLIDLIARFKTOGG 733
Oy      796 YIWSDF--NHVYRLRFAGDOSIRGYAHDLSLPSIDKGYLGGQVLAVGTAEVNYEFM 854
Db      734 YIFRYNTDYLPLANSFTFYMGQVTVYVFRNGSLTPKDEFGLMGDCIFPASTELSGVL 793
Oy      855 K--DLRLAVFGDIG-----NAYDKGFTNDT-KIGAGGV 885
Db      794 KAKMKRLAFEFDFGLFTKPTRGSEFYNAPTTANFKDYGAGGERATWASGLQI 853
Oy      886 RNASPYG 892
Db      854 EWISPMG 860

RESULT 36
O51930 PRELIMINARY: PRT: 789 AA.
AC O51930: O51921:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE Outer membrane antigen OMA87.
GN OMA87.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxId=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PBA100;
RA Ruffolo G.C., Adler B.;
RL Infect. Immun. 64:0-0(0).
RN [2]
RP SEQUENCE OF 750-789 FROM N.A.
RC STRAIN=9222;
RX MEDLINE=95369730; PubMed=7642134;
RA Delamarche C., Manoha F., Behar G., Houlgatte R., Hellman U.,
RA Wroblewski H.;
RT "Characterization of the Pasteurella multocida skp and fira genes.";
RL Gene 161:39-43(1995).
DR EMBL: U60439; AAC44600.1; -
DR InterPro: IPR00184; Bac_surfaG_D15.
DR InterPro: IPR001064; Crystallin-
DR Pfam: PF01103; Bac_surface_Ag_1.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Hypothetical protein.
```

```

SQ      789 AA: 87437 MW: DCID004AAEB2CD9E CRC64;
Query Match 4.1%; Score 194; DB 2; Length 789;
Best Local Similarity 19.0%; Pred. No. 0.0058;
Matches 154; Conservative 125; Mismatches 304; Indels 228; Gaps 37;

Oy      247 YVDIDLSIIRNSIGEDVVIHDLGEPYIDYRAVEVBEGADKRAFTVADEVPLIGDV 306
Db      69 YDDVKASREGNTL-----VTVMPKPVISN--VYIVGNKSLPDEAIKQNDANGFKVGDV 121
Oy      307 FHHGKYET--KKNLIENASAEHGFDGKRWLDRSDVTL--LPDNTADVSLIYDTGTQYRFD 362
Db      122 LMRKLEEFRRKGIYEHVNSV-----GRY--NAKYDAIYNTLPPNSAEIKIQINEDDVALFK 175
Oy      363 EYVF-----FTIDPKTNO--LTTDP-----DKLPVRELLEOLLYVMGEAYNLQA 406
Db      176 EITFEENAEASSGKLADOMELQDSMKKLGKNGKFDQYQFNKD--LETIRSYLDRGY----- 230
Oy      407 VRALSNDLIATRYFNMYNTEIVEPERQIONDOVSFBOSSSRTEPAQVDE----- 457
Db      231 -----AQFQLDITV-----KLSDDKKEPCLISEGDLVYTKT 263
Oy      458 -----STLEPIETVELTDGILMDISPIESASMLIQDKLN-----LYAAARH 501
Db      264 RVSGMGMGMSAEIAPLLETIQL--NGLFRTSYLEVQQRN--KSKLGERGYATAQVAVHP 320
Oy      502 LYDMPDORVLAIHNDGVNSIL-----GRISDAVSAVARAILPDES--ENEVIDLP 551
Db      321 TFD--EODKTTSLDPIYBAGKSYVROIRFEGNNTSSADSTLRQERQOEGMLSELVELG 379
Oy      552 -----ERTAL--ANRKPADYQSKKVPYLVFVASDKPRDQIGLGSDGTGRLVYKF- 603
Db      380 KLRIDRTGYEVSVEYTKEA--IPGSDQVDVLYKKERTGSLINGIGTSGSLYQASIK 438
Oy      604 EHNLI-----NDGYQAGAEHLRSL--DKKGVLYATKPLSHPLNDQLRATLG 650
Db      439 QDNFLMGSSISLGTRNDYGTYNLGYNEPYFTKQDVSIGWVSFEEDYSSKNSAGY 498
Oy      651 QOEYF--GHSTNGFDLSRTLHEHSRIIONGMNRTYSL--RYRLDKLTKTOPP----- 702
Db      499 GRTSYGGLTLGFPVNV-----NNSYLYGVGYNKLNLNIAPEYNRDL 541
Oy      703 -----ETMODLPVDFVNGKPSQEALLAGVAVHKTVAADNLVPMRGYQRYRSLEVSS 754
Db      542 YKOSMKYNDWTFKSHDF-----DLDFGNMYNSLNGRYPTPGVARNIG 585
Oy      755 GLV----SDANMAIARAGISGVYSPFDNAYGSNRAHQMTGIGIAGYIWSDNFN--HVPYR 808
Db      586 GRVTIPGSDNKKYKLAEAQGFYPL-----DREHGWLSSRISASFADGFGSKRLPFY 638
Oy      809 LRFAGGDSIRGYAHDLSLP-----ISDKGLTGGQVLAVGTAE 848
Db      639 QYTSAGIGISLRGFAVGALGPNAIYRTRQCPDSYCLVSSDVIGNMAVYASTELIYPTP 698
Oy      849 YNTEFMKDLRLAVFGDIGNAYDKGFTNDTK-----IGAGVWRHASP 890
Db      699 VAKKNQNSVYRSLFVDAASAVWNRKAKEDKAKFAKLANVPYSDPSKVRASAGVALQWOSP 758
Oy      891 VGOVRVAVATGKE--EGNPI--KLHFIIGTF 919
Db      759 IGPLVSYAKPLKKYQGDIEQFQSIGTF 789

RESULT 39
O8YMP0 PRELIMINARY: PRT: 676 AA.
AC O8YMP0:
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Hypothetical protein Alr4893.
GN Alr4893.
OS Anabaena sp. (strain PCC 7120).
```

```

OC Bacteria: Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
RX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT Cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
DR EMBL: AP003597; BAB76592.1; -.
DR InterPro: IPR000977; DNA_Ligase.
DR PROSITE: PS00697; DNA_LIGASE.A1; UNKNOWN_1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 676 AA; 72411 MW; C577963037D1223 CRC64;

Query Match 4.1%; Score 193.5; DB 16; Length 676;
Best Local Similarity 20.1%; Pred. No. 0.0048;
Matches 191; Conservative 119; Mismatches 323; Indels 317; Gaps 47;

OY 13 MPVALAAYLRLMSQALAQONNPANIINHVPADHTAINQAKGNPVLTPPEQIARLNA 72
DB 1 MRISAAALLTLTLVA-----SNATQAIAA-PHTASTPSK----- 35
OY 73 AGLNARQSALQVNVNDDSPISRIGESPPRLGLDMSEVEETPLLELFAQESTEMG 132
DB 36 -----QENNL-VIRVVEETPV-QLGAVTPP-----ET-----EVTFQFSTKSR 71
OY 133 INPDVITYEYQGEQNPSEVVVPPTLEPEKPLIKRLARLFND-----GVNKPRL 163
DB 72 VIAQ-----NSPVLPSTPRPVPPAKP--ATTVNLVYATDVQVIGATPEL 118
OY 184 K---AKFYSSQSGETSALIGSSHQKTEPYANIKALEDTIQESAMDLNGSIPRLROTALV 240
DB 119 QETIRQVINTQGTGDS-----QTQLRDVAALIE-----TGLE 152
OY 241 AARAVGYUDIDLSIINSIGEVYIIHDLGEPYIDYRAVENEGEADDAKAF--TYADEV 299
DB 153 ASANVN-----SKTTPSGLVNVVO--VQPVV--RSLQLGCA---KALVSYVAQ-- 194
OY 300 PLLIGDVFNHGKYEKKNL-----IENASAEHGYPFDGRML--DRGVYLIPDNTADVSL 351
DB 195 PRFQSOI---GPRISPEGLKQAVAOVMYADNGVNLAVLSLEPRQSLININVAE--GL 250
OY 352 IYDTQYVFEDEVFFETIDPKTNQL--TTDPKLPVKKRELLQQLLTVMNGEAVNLQAVRA 409
DB 251 VSD--IKFEFVNDGDKTIDNSNGNPVGGRTKPD-----FLRQQLKLGQGVFOENTVQ 301
OY 410 LSNDLIATYFNMVTEIYFPEREQIONQVSFEQSSSRTEPAOVDESTLEPVLETVEL 469
DB 302 DVQQLYRGLFGOSV----- 316
OY 470 TDGILMDISPIEFSASNLIDKLNVAAKARHLVMDPDRVLAINHDDGVNSRI--LGR 527
DB 317 -----VAFAGD-----ATKLDIYELKENGARAINLIGSSYNDDVGLMGT 356
OY 528 SDVAVAVARAILPDESENEVIDLPERTALANRKTADVQSKKVPVLYVAVASDKPRDQ 587
DB 357 N-----YQDNIIGKNDTLIAN-----V 374
OY 588 GLGMSDGTGRLYTKF--EHNILNRD--GYQAGAFRLSEDKKGYKLYATKPLSPMDQ 643
DB 375 GL-----SRDLQFDTKETISPYDPTNSDRGLGYVNA-----FRREISSETPDE 418
OY 644 LRATLGYQOEVEGSHSTNGFDLSTRTLEHISRSIIQNGCMNRYSLSVRKDKIKTQ--A 700
DB 419 IKLANGKQVRE--GKVGIGISIQ--RPID-----GMAISIGFNYSRISINDROGNI 465
OY 701 PPEWMDLPVDFVNGKPSQELLAGVAVHKTVDNLVNPMKYRQRYSLVEGSSGLVSDA 760
DB 701 PPEWMDLPVDFVNGKPSQELLAGVAVHKTVDNLVNPMKYRQRYSLVEGSSGLVSDA 760

OC Bacteria: Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
RX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Ulfenback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005865; AAK23890.1; -.
DR TIGR: CC1915; -.
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
DR Complete proteome.
SQ SEQUENCE 769 AA; 85938 MW; 7B12A1FEB6226F2B CRC64;

Query Match 4.0%; Score 190.5; DB 16; Length 769;
Best Local Similarity 20.0%; Pred. No. 0.009;
Matches 151; Conservative 94; Mismatches 250; Indels 259; Gaps 34;

OY 250 IDLSIIRNSIGEVYI--IHDLGEPYI--DYRAVEYRGSGADDAKFTYADVPELLIGDV 306
DB 160 VDLVEINEGAKSGVGLNFDLGNAEYSNDLRDVIYTKESRYKILTS-----NDN 210
OY 307 FHHGKYEKKNLIEENASAEHGYPFDGRMLDRSDVVLIPD--NTADVSLIYDTQYREDEV 365
DB 211 YDPRILEYDRQGLRKHYNRGFFDFRYS--SVALAPDKNGAVYTTLEEGKTKAFGR 268
OY 366 FFTIDPKTNQLTTPDKLPVKKRELLQQLLTVMNGEAVNLQAVALSNDLIATRYFNMYNT 425
DB 269 --TYETELKILDGN-----LLAQILLPRTGLYE----- 295
OY 426 EIVPEREQIONQVSFEQSSSS--RTEPAQVDESTLEPVLETVELTGDGLMDISPIE 481
DB 296 -----DERIEQATDALTFAAGAGAFVDVPRPYPNRETGYVDVVFQVREG----- 342
OY 482 FSASNLIDKLNVAAKARHLVMDPDRVLAINHDDGVNRSILIGISDVAVAVARAILPD 541
DB 343 ---PRVYVDRIIDY--GNRTLL--DYVLRRELEVAEGDAVNR-----VLYVD 381
OY 542 ESEN-----EVID-----LPERTALANRKTADVQSKKVPVLYVAVASDKPRDQ 586

```



```
Db 382 RSKNNRRRLGFEFEVEDAPGSA PDRTSLR-----VKVEEQPTGE 422
QY 587 IGLWGS D T G T R L V T F E H N L I N R D G Y O A G A E I R L S E D K K G V K L Y A T K P L S H P L N D Q L R A 646
Db 423 L S F S A G Y S S I D K L V - L D V G I T E R N F R G R G Q N L R-----ARA 457
QY 647 T L G - Y O Q E V F G H S T N G F D L S T R T L E H E I R S T I O N G M N R T Y S L R Y R L D K L T Q A P P E T 704
Db 458 S V G S L R Q O I D F G F S E P R F-----I G R N L Y - A G V N - L Y T F R Y D L S E F A Y D T K S V 504
QY 705 M O D L P V D F V N G R K P S Q E A L L A G V A V H K - T V A D N L-----736
Db 505 G G D V R G F P L T N D S M S L R T V R O D E V S A D S L C A S G S V S Q I L C L O R A Y I T S L I G Y G L R 564
QY 737 -----V N P M R G Y R O R Y S L E V G S S G L V S D A N M A I A R A G I S G V Y S F G D N A Y G S N R A-----785
Db 565 I D K R N D P I N F T R G W-----F A D L N Q D L A G V-----G D V K Y L K T E A D A G W 604
QY 786 -----H O M T G G I O A G I Y M S D N F N H P Y R L R F F A G D Q S I R G Y A H D S L S P--I S D K G 834
Db 605 Y M G F T R D L V F S A T G S F--G Y I E G M G D N V R I N D R F Y R G G--T S F R G F E I A G I G P R D I S S F 661
QY 835 Y L T G G O V L A N G T A E Y N Y - E F M K D--L R L A V F G D I G N A Y D K G F T N D P-----877
Db 662 N S M G A K L Y A I S T F E L V P T L P E Q Y G I K A L F S D V G T A--G L L D V D R O R S P G V F D P N I 718
QY 878 -----K I G A G V G R M A S P V G O V R V D A T G V K E E 905
Db 719 K D N L G L R A S A G I S I D M K S P M G P I R F D I S R I L S K E 752
```

Search completed: April 28, 2003, 16:23:06

Job time : 68 secs

